DIVERSE ROLES OF NUCLEAR INTERMEDIATE FILAMENTS IN PROLIFERATING CELLS

Manasvita Vashisth

A DISSERTATION

in

Chemical and Biomolecular Engineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

____ _____

Dennis E. Discher

Robert D. Bent chaired Professor of Chemical and Biomolecular Engineering

Graduate Group Chairperson

_____ _____

John C. Crocker, Professor of Chemical and Biomolecular Engineering

Dissertation Committee

Rebecca G. Wells, Professor of Medicine

Ravi Radhakrishnan, Professor of Bioengineering

John C. Crocker, Professor of Chemical and Biomolecular Engineering

DIVERSE ROLES OF NUCLEAR INTERMEDIATE FILAMENTS IN PROLIFERATING CELLS

Dedicated to my parents

Ajay & Monika Tyagi

ACKNOWLEGMENTS

I would like to thank Prof Dennis Discher for all the informative discussions and taking time out of your schedule to give very detailed feedback on an almost daily basis. I had read about your work in India and it had been a dream come true to be able work with you. It has been an honor and a privilege.

I also want to thank all the members of the Discher lab, past and present, especially Sangkyun Cho, Charlotte Pfeifer, Manu Tewari and Jerome Irianto, for having the patience to teach me all the meticulous wet lab work.

I also would like to thank my PhD cohort and friends from undergrad for providing me with mental strength and support on this journey that can otherwise be very solitary. Thank you to my cousins who provided me a home away from home.

Last and most importantly I want to thank my parents for being my pillars of strength, I could not have done this without you.

ABSTRACT

DIVERSE ROLES OF NUCLEAR INTERMEDIATE FILAMENTS IN PROLIFERATING CELLS

Manasvita Vashisth

Dennis E. Discher

Embryonic tissues and cancer have in common the fact that they are both highly proliferative tissues rapidly moving through the cell cycle, as opposed to most other differentiated tissues in an adult. DNA damage can arrest some embryonic cells but genetic instability is a hallmark of cancer. This thesis studies the contrasting role of two nuclear intermediate filaments - Lamin A and Lamin B1 in the proliferating cells of embryonic hearts and cancer. Lamin B1 is upregulated together with proliferation genes in at least 15 cancers curated in The Cancer Genome Atlas (TCGA), whereas Lamin A trends align with 'matrix mechanosensititve' genes. With physicochemical principles in mind, we show Lamin B1 scales with many mitosis genes in cancer, and experiments reveal its role in promoting cell cycle and direct regulation by the cell cycle transcription factor FOXM1. The genes that scale are used in Scaling-informed Machine Learning (SIML) to better predict overall patient survival and to better identify cell lineage in single cell RNA profiles. A distinct role of Lamin A is revealed by experiments on the first organ in its first days – the heart – which show Lamin A levels are modulated in interphase cells through phosphorylation in response to acto-myosin stress. Lamin A levels determine the probability of nuclear rupture and subsequent DNA damage, telomere attrition, and cell cycle arrest. Nuclear lamins thus have different roles in responding to and regulating cell cycle.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1 Introduction

Embryonic/developing tissues and cancer have in common the fact that they are both highly proliferative tissues rapidly moving through the cell cycle, as opposed to most other differentiated tissues in adult human. However, there are key difference in terms of regulation of proliferation in development and cancer. Heart disease and cancer combined caused 43.2% of the deaths in the US in 2019 (Kochanek, Xu and Arias, 2019), however their underlying pathophysiology has contrasting traits. The basis of heart failure can be attributed to the failure of the myocardium to regenerate itself, proliferative ability of cardiomyocytes drops sharply shortly after birth(Porrello *et al.*, 2011, 2013). DNA damage including telomere defects, micronuclei etc arrest cardiomyocytes, in sharp contrast, chromosomal instability and DNA damage promote cancer development. This thesis studies the role of two nuclear intermediate filaments- Lamin A and Lamin B1 in the proliferating cells of embryonic hearts and cancer. Lamin A is mechanosensitive which plays a role in determining the probability of nuclear membrane rupture and subsequent accumulated DNA damage and cell cycle arrest. Lamin B1 in cancer cells acts as a functional marker for proliferation, regulated by FOXM1.

1.1 Properties of Lamins

Below the inner nuclear membrane exists a meshwork of proteins called the lamina that confers mechanical support and stiffness to the nuclear envelope(Ungricht and Kutay, 2017). Type V intermediate filament proteins called lamins form the majority of the lamina. Lamins undergo self-assembly into filaments of protein that form the meshwork. We focus our study on two major groups of lamins- A type lamins and B type lamins.

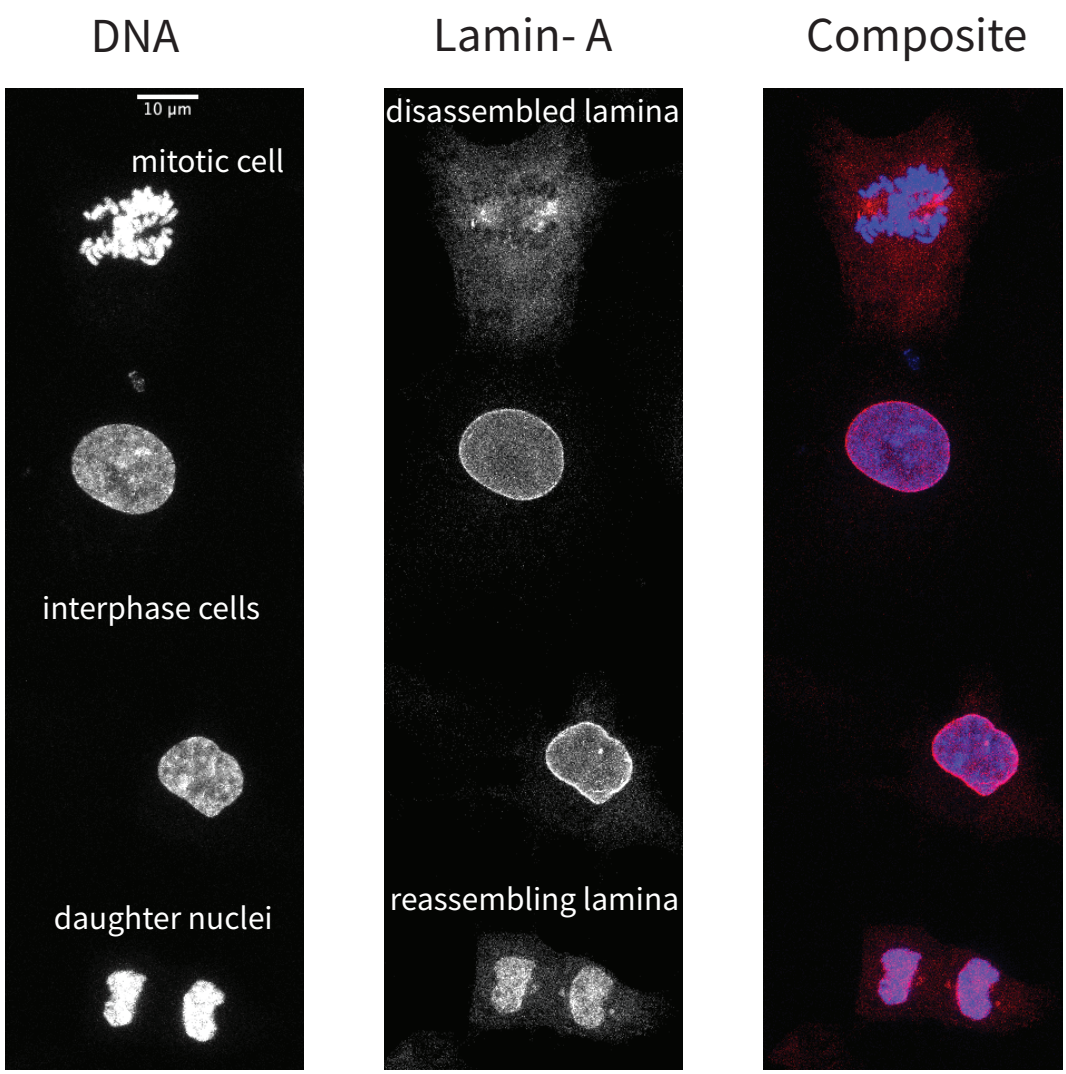


Figure 1-1 Lamina disassembles during mitosis and reassembles in daughter cells
63x magnified confocal images of lung adenocarcinoma epithelial cells-A549 with GFP tagged Lamin-A and Hoechst 33342 stained DNA

1.1.1 Evolution

Intermediate filament proteins evolved later than other cytoskeletal proteins like actin. Evidence suggests that B type lamins evolved before A type lamins given that Lamin B1 is constitutively expressed in all invertebrates unlike Lamin A(Dittmer and Misteli, 2011; Peter and Stick, 2012).

1.1.2 Structure

Lamins have a nuclear localization signal (NLS), an immunoglobulin (Ig)-fold domain, and a CaaX motif at the C terminus(Gruenbaum and Aebi, 2014). There are predominantly four lamin isoforms in mammalian cells- two A-type lamins (LMNA and LMNC) and two B-type lamins (LMNB1 and LMNB2). Different lamins undergo different post translational modifications (PTMs) resulting in different biochemical and physical properties. The CaaX motif at the C terminus of LMNA, LMNB1 and LMNB2 is farnesylated which is later proteolytically cleaved for LMNA(Hennekes and Nigg, 1994). This results in only B-type lamins being permanently anchored to the INM, even during mitosis (in the form of fragments)(Gerace and Blobel, 1980).

1.1.3 Mitosis

The lamina, along with the nuclear membrane, reversibly dissociates (sometimes to monomeric units) at the beginning of mitosis and reassembles in the daughter cells (Gerace and Blobel, 1980). They are disassembled before mitosis by phosphorylation by specific kinases-CDK and PKC (Dechat *et al.*, 2010) and reassembled beginning of interphase of daughter cells through dephosphorylation by phosphatases (**Fig 1.1**). This reversible phosphorylation and depolymerization of lamin proteins into a soluble state has also observed in interphase cells. We further study the role of this solubilization flux in interphase cells in Chapter 2. However, there are differences in the B-type lamins and and A-type lamins. The B-type lamins remain farnesylated to the lipids of the nuclear membrane. Lamin B1 accumulates at the chromatin during the anaphase-telophase transition. In contrast, lamin-A begins polymerizing at the lamina much later and remains diffuse in the cytoplasm even during G1 phase of the cell cycle of the daughter cells(Moir, Yoon, *et al.*, 2000). These differences in reassembly are illustrated Chapter 3 that shows *LMNB1* in primary tumor tissues universally scales with mitosis associated genes like kinesins, unlike *LMNA*.

1.1.4 Expression in healthy tissues

Lamin-A expression increases as the stiffness of the resident tissue increases whereas Lamin B1 is constitutively expressed and does not scale with stiffness (Swift *et al.*, 2013). Embryonic tissues

stiffen overall as the embryo develops, Lamin B1 is high from the beginning and remains constant as the embryo develops whereas Lamin A is almost undetectable at the beginning but responds to the stiffening tissues and increases with time, indeed it is one the proteins that show the highest fold change as the embryo develops(S. Cho *et al.*, 2019). In Chapter 2 we study the role of upregulation of Lamin A as the embryo develops.

1.1.5 Rheological properties

Nuclei have viscoelastic properties(Guilak, Tedrow and Burgkart, 2000; Dahl *et al.*, 2005), meaning that they exhibit stress relaxation: when a constant deformation is applied, the resulting mechanical stress on the nucleus decays over time. Micropipette aspiration experiments reveal that Lamin-A and Lamin B have different rheological properties with Lamin A being more viscous whereas Lamin B1 is more elastic(Swift *et al.*, 2013). This can have consequences for constricted migration of cells. Constricted migration of cells is a physiologically important phenomena for many process in the healthy tissue like -embryo development(Kurosaka and Kashina, 2008), angiogenesis(Lamalice, Le Boeuf and Huot, 2007) and wound healing(Clark *et al.*, 1982) as well as in malignant tissue - cancer metastasis (Liotta, Steeg and Stetler-Stevenson, 1991). The nucleus being the largest and stiffest cell organelle acts as a rate limiting factor in constricted migration. As the nucleus moves through tiny pores, the nuclear lamina can develop blebs (**Fig 1.2**). Some blebs can be Lamin B1 negative since Lamin B1 is elastic and cannot sustain shear stress, whereas Lamin A being more viscous, ‘creeps’ into the blebbs(Pfeifer *et al.*, 2018; Xia, Pfeifer, *et al.*, 2018).

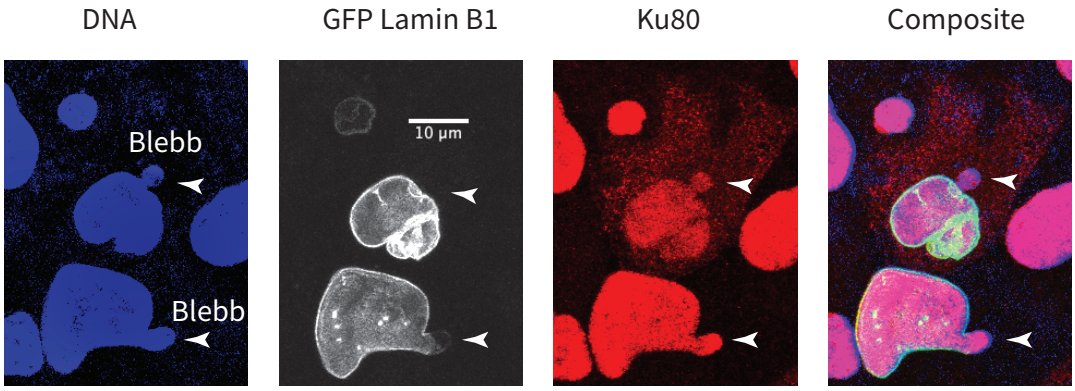


Figure 1-2 Lamin B1 negative blebb in nuclei after undergoing constrictive migration

Confocal images of transfected GFP-LMNB1 osteosarcoma epithelial cells U2OS after undergoing constrictive migration through 3m transwell pores, immunostained with nuclear protein- Ku80, DNA stained with Hoechst 33342

1.1.6 Mutation

Diseases arising due to Lamin B1 mutations are rare, perhaps hinting at functional Lamin B1 being essential for early development of embryo. In contrast, Lamin A mutations are associated with many diseases – skeletal/cardiac muscles diseases (Emery Drefuss Muscular dystrophy (EMD), dilated cardiomyopathy (DCM), limb girdle muscular dystrophy (LGMD1B), redistribution of white fat (familial partial lipodystrophy (FPLD), mandibuloacral dysplasia (MAD)), charcot marietooth neuropathy (CMT2) and progeria (accelerated aging).

1.1.7 Senescence

Lamin B1 is drastically downregulated in senescent cells and a well-established marker of senescence(Shimi *et al.*, 2011; Freund, R.-M. Laberge, *et al.*, 2012; Shah *et al.*, 2013), independent of Lamin A.

1.2 Functions of Lamins

Lamins have diverse roles including maintaining nuclear size and shape, mechanosensing, genomic organization, regulating transcription etc. Here we illustrate a few of them.

1.2.1 Maintaining compartmentalization in a cell

Nuclear membrane plays the important role of separating the nucleoplasm from the cytoplasm to maintain the integrity of DNA, keeping DNA out of the way from dynamic events like trafficking along the cytoskeleton, compartmentalization of transcription and splicing steps away from translation, spatially organizing DNA metabolism events for efficiency and maintaining the concentration of nuclear proteins like DNA repair factors, transcription factors, spatially regulating transcription factors or signal transduction networks by modulating their nuclear to cytoplasmic ratio. However, all proteins are translated by ribosomes in the cytoplasm/endoplasmic reticulum. Proteins with molecular weight >~45 kDa need to be transported by the importin superfamily across the nuclear pore complexes (NPC) - 'gates' of nuclear membrane. Proteins that need to end up in the nucleus have a 'nuclear localization signal' (NLS) which is a sequence of amino acids that is recognized by an adaptor protein called Importin α which dimerizes with Importin β, which mediates interactions with the NPC. The analogous process involves the exportin protein family transporting proteins with a Nuclear Export Sequence (NES) out of the nuclear pore complex and into the cytoplasm, for example to transport the transcribed mRNA, tRNA, ribosomes and other proteins. Different proteins can have a wide variety of NLS/NES, moreover, there are 7 Importin αs and >20 Importin βs(Wagstaff *et al.*, 2012). Some proteins can have both NLS and NES. There are two major types of NLS – monopartite (one continuous sequence of amino acids, for example the one present in SV40 virus) and bipartite (NLS present as two clusters of amino acids). For example, Ku heterodimer consisting of Ku80 (*XRCC5*) and Ku70 (*XRCC6*) play a central role in DNA repair. Anti-parasitic drug Ivermectin acts as a specific inhibitor of importin α/β mediated nuclear transport. Each of the monomers can independently transport to the nucleus through their respective monopartite NLS as well as heterodimerize for nuclear transport(Koike, Shiomi and Koike, 2001; Takeda *et al.*, 2011). Exportin chromosome region maintenance 1 (CRM1) is the most common and exports many proteins and is inhibited by antifungal drug Leptomycin(Wagstaff *et al.*, 2012).

Other exportins include Exportin-t that directly binds to tRNA to transport outside the nucleus(Arts *et al.*, 1998).

Nuclear membrane rupture

This important and regulated compartmentalization by the nuclear membrane can be voided in cells undergoing constricted migration(Denais *et al.*, 2016; Irianto, Charlotte R Pfeifer, *et al.*, 2016; Irianto *et al.*, 2017) and repair of the rupture through ESCORT III recruitment can be critical for subsequent cell proliferation due to accumulation of DNA damage(Raab *et al.*, 2016). Interestingly, it is not the physical pulling apart of the DNA as the cell moves through confined spaces that causes increase in DNA damage since the tightly packed DNA unravels as it experiences shear stress(Irianto, Xia, *et al.*, 2016) but the mis localization of DNA repair factors. Lamin A provides stiffness to the nucleus and plays an important role in constricted migration, Lamin A knockdown cells, migrate faster but are also more likely to undergo rupture which limits their survival (Harada *et al.*, 2014a). Moreover, cancer cells on stiff plastic 2-D matrix can also exhibit nuclear membrane rupture and prolonged mis localization of DNA repair factors leads to eventual accumulation of DNA damage and cell cycle arrest(Xia, Ivanovska, *et al.*, 2018). In chapter 2 and 4 we further expand this study by using nuclear import inhibitors specific to DNA repair proteins to verify this hypothesis.

DNA damage induced cell cycle arrest can have biological significance as post-natal hearts permanently exit cell cycle shortly after birth, which has previously also been attributed to DNA damage due to oxidative stress(Puente *et al.*, 2014). Interestingly Ku80 deficient mice show progeroid symptoms as well as early senescence similar to the disease progeria caused by Lamin A mutations (Vogel *et al.*, 1999; Reiling *et al.*, 2014). In chapter 2, we posit that increasingly stiff extra-cellular matrix (Sangkyun Cho, Vashisth, Abbas, Majkut, Vogel, Xia, Irena L. Ivanovska, *et al.*, 2019) and acto-myosin stress in a developing embryo with deficient Lamin A increases the probability of nuclear rupture and accumulation of DNA damage, telomere attrition and cell cycle arrest.

1.2.2 Cancer progression and metastasis

Dysmorphic nuclei are a hallmark of cancer and lamin levels are often dysregulated(Irianto, Charlotte R. Pfeifer, *et al.*, 2016) but there is no consistent trend towards up or down regulation that pervades all cancers. Reduced expression of Lamin A in some malignant cancers may be resultant from an adaptive response to soften the nucleus for ease of constrictive migration. Lamin B1 is upregulated in many cancers with circulating Lamin B1 acting as a prognostic marker in Liver cancer (Sun *et al.*, 2010; Abdelghany *et al.*, 2018). In chapter 3, we strive to find 'universal' scaling laws that pervade across many cancer types and reveal transcriptional regulator of Lamin B1.

1.2.3 Regulation of DNA transcription and replication

Chromatin has various sub-structures through DNA being wrapped around histone octamers and exists in two forms: gene-rich euchromatin (low density, perhaps for easy access to transcription factors and RNA polymerase); and gene-poor heterochromatin (higher density). There are regions of the primarily heterochromatin that directly bind to the lamina called Lamin associated domains (LADs). LADs can change as an embryo develop and play a role in transcriptional regulation. Moreover, cell culture studies show Lamin-B1 depletion inhibits DNA replication and Lamin B1 foci n the nucleoplasm coincide with sites of DNA replication (Moir, Spann, *et al.*, 2000)-(Moir, Montag-Lowy and Goldman, 1994). Moreover, Lamin-B1 knockdown undermines early development of embryos, unlike Lamin-A (Harborth *et al.*, 2001a). Chapter 3 shows Lamin B1 increases as the cell proceeds through the cell cycle and knockdown of Lamin B1 can reduce the number of cells that proceed through the cell cycle. However, the exact role of Lamin B1 in DNA replication remains unclear.

1.3 Conclusion and outline

Even though different types of lamins have a largely similar structure, they can have vastly different properties and functions. In Chapter 2, we first start with studying what is the role of Lamin-A upregulation in developing embryos. Lamin A mutations underly the disease of accelerated aging-

progeria, concurrently Ku80 deficient mice show similar phenotype. We strive to find the role of Lamin A in aging and telomere maintenance. This can be contrasted to the lack of modulation and constitutive expression of Lamin B1 (except in senescent cells), perhaps hinting at its necessity in proliferating cells. This leads us to study Lamin B1 expression in cancer patients to find that *LMNB1* is upregulated in the tumor tissue across 15 different cancers, moreover higher expression of Lamin B1 significantly predicts poor survival in patients across 8 cancers, unlike *LMNA* that does not show consistent upregulation across cancers.

Each cancer can have a very specific pathophysiology, indeed it is hard to find a 'cure-all' panacea for cancer, given each type of cancer is a very different disease altogether. As personalized medicine emerges and more and more patients get their genome sequenced, we find that within a cancer type also patients can have a multitude of different malignancies which evolve with time. There have been efforts by a massive collaboration through the National Cancer Institute (NCI) which formed The Cancer Genome Atlas to combine data from over 9112 patients across 32 cancer types in a pan-cancer atlas.

In chapter 3, we aim to find 'universal' scaling laws that span across cancers and focus on expression of Lamin B1 based on its consistent upregulation across cancers. New methods in machine learning for dimensional reduction and visualization are unable to cluster patient transcriptomes based on patient survival. We use 'universal' scaling laws to make meaningful inferences of transcriptional regulation, patient survival prediction and identifying cell type markers in tumor tissue.

One of the differences between normal cells and cancer cells is illustrated by the Hayflick limit which states that normal cells in culture can double only approximately 50 times, in contrast, cancer cells have no Hayflick limit(Hayflick and Moorhead, 1961). Cancer is a disease of old age that accumulates environmental stress as well as genetic pre-disposition, as you grow older, more and more mutations accumulate, even if you are genetically pre-disposed by a mutation, the cells have redundancies and at least two or more tumor suppressor genes need to be silenced for malignancy

to emerge as per the Knudson two hit hypothesis (Knudson, 1971). Hence, DNA damage to cells is an important factor in the emergence of cancer and genomic instability is a hallmark of cancer.

Constitutively expressed Lamin B1 can be tagged with a fluorescent marker to track genomic changes in live cancer cells. We develop a pipeline for big data analysis of single cell sequencing which allows us to link genomic changes to mRNA sequencing profiles of cells and determine the lineage of genetically unstable cancer cells.

Low levels of genetic instability and aneuploidy have also been reported in mammalian embryos that have sustained genetic mutations(Eggan *et al.*, 2002; Knouse *et al.*, 2014; Singla *et al.*, 2020). We aim to use the pipeline developed above to study if the DNA damage sustained in embryonic hearts illustrated in Chapter 2, can lead to genetic rearrangements. This would attempt to rationalize the loss of regenerative capacity in post-natal hearts as perhaps a protective mechanism against cancer development, given that heart tumors are vanishingly rare.

Chapter 2 Mechanosensing by Lamin A protects against nuclear rupture, DNA Damage, cell cycle arrest and aging

This chapter has been adapted from the published paper Cho S., Vashisth M., et al (2019) 'Mechanosensing by the Lamina Protects against Nuclear Rupture, DNA Damage, and Cell-Cycle Arrest', Developmental Cell, 49(6). doi: 10.1016/j.devcel.2019.04.020

Computational analysis in Max Quant Perseus of the mass spectrometry data in Fig 2.4d,e has been done by Sangkyun Cho. Mass spectrometry was performed at the Wistar Institute. hiPS-cardiomyocyte organoid perturbations were done by Manu Tewari. Chicken embryonic hearts extraction was done by me in conjunction with Manu Tewari. Telomere Q-FISH experiments in Fig 2.4i were done by Elisia Tichy.

2.1 Abstract

Whether cell forces can impact genome integrity is largely unclear. Here, acute perturbations (~1hr) to actomyosin stress cause rapid and reversible changes in lamin-A, DNA damage, and cell cycle. The findings are especially relevant to organs such as heart because DNA damage permanently arrests cardiomyocyte proliferation shortly after birth and thereby eliminates regeneration after subsequent injuries including heart attack. Concurrently, effects are also observed in proliferating immortalized cancer cell lines. Embryonic hearts, immortalized lung adenocarcinoma epithelial cells (A549) and immortalized osteosarcomia epithelial cells (U2OS) all show that actomyosin-driven nuclear rupture causes cytoplasmic mis-localization of DNA repair factors and accumulation of DNA damage leading to eventual cell cycle arrest. Mislocalization of NHEJ DNA repair pathway protein Ku80 is shown to produce DNA damage specific to telomeres. Lamin A knockdown shows phenotype of increased DNA damage and cell cycle arrest similar to knockdown of DNA repair factors. This is mechanistically caused by increased frequency of nuclear membrane rupture when Lamin A is low but actomyosin stress is high. Lamin-A is thus stress-stabilized to mechano-protect the genome.

a2.2 Introduction

Proliferation of many cell types slows dramatically shortly after birth and is absent in key adult tissues(Li *et al.*, 1996), which presents a major challenge to regeneration. Cell growth in culture is modulated by the stiffness of extracellular matrix (ECM) which generally promotes actomyosin stress (Engler *et al.*, 2006) and affects the conformation (Sawada *et al.*, 2006), post-translational modification (PTM) (Guilluy *et al.*, 2014) , localization (Dupont *et al.*, 2011), and degradation (Dupont *et al.*, 2011) of mechanosensitive proteins. Actomyosin links to ECM and to nuclei, such that rapid changes to ECM and tissue mechanics that result from chronic or acute injury or even drug therapies (e.g. cardiac arrest or cardiomyopathy treatment can in principle affect the nucleus (Kanellos *et al.*, 2015; Takaki *et al.*, 2017) and perhaps the DNA within. DNA damage and telomere shortening are indeed well-documented in injuries that affect heart(Oh *et al.*, 2003; Higo *et al.*, 2017; Sharifi-Sanjani *et al.*, 2017; Chang *et al.*, 2018) as well as nonmuscle tissue – but DNA damage and repair are rarely studied in developing organs, and relationships to proliferation, ECM stiffness, and actomyosin stress are understudied.

DNA damage and senescence increase with many disease-linked mutations, including those in the nucleoskeletal protein lamin-A (LMNA) that forms a structural meshwork around chromatin (Shimi *et al.*, 2008; Gruenbaum and Aebi, 2014; Turgay *et al.*, 2017). LMNA deficiencies associate with elevated DNA damage (Graziano, et al., 2018; Liu, et al., 2005) and result in accelerated aging of stiff tissues similar to deficiencies in DNA repair factors (e.g. KU80) (Li, et al., 2007). Moreover, progeroid syndromes are caused only by mutations in LMNA and DNA repair factors, but LMNA's primary function in development remains hotly debated (Burke and Stewart, 2013), with suggested roles in gene positioning and regulation (Harr, et al., 2015) seeming at odds with largely normal development of human and mouse mutants until weeks after birth. Surprisingly, senescence or apoptosis of cells with LMNA defects is rescued by culturing cells on almost any ECM (versus rigid plastic (de La Rosa, et al., 2013; Hernandez, et al., 2010)) and by treatment with at least one drug affecting both cytoskeleton and nucleo-cytoplasmic trafficking (Larrieu, et al., 2018; Larrieu, et al.,

2014). Relationships between lamins, actomyosin stress, ECM mechanics, and DNA damage are nonetheless obscure – especially in tissues.

Embryonic hearts beat spontaneously for days after isolation from early chick embryos, and beating is acutely sensitive to myosin-II inhibition as well as enzymatic stiffening or softening of ECM (Majkut, et al., 2013). DNA damage is conceivably optimized in heart as it triggers a switch from proliferation to senescence in post-natal hearts (Puente, et al., 2014). DNA damage is also implicated in telomere attrition and binucleation of CMs that signal irreversible exit from cell cycle (Aix, et al., 2016). We postulated embryonic hearts with rapidly tunable mechanics could prove useful as a tissue model for clarifying protein-level mechanosensing mechanisms in vivo that could be studied thoroughly in vitro with many cell types.

Although LMNA is reportedly 'undetectable' in early embryos (Stewart and Burke, 1987), it progressively accumulates in tissues such as heart, bone, and lung (Solovei, et al., 2013; Rober, et al., 1989; Lehner, et al., 1987) and is highest in adults within these mechanically stressed, stiff tissues that are collagen-rich (Swift, et al., 2013). LMNA deficiencies accordingly produce measurable defects weeks after birth in such tissues, including heart (Kubben, et al., 2011; Worman and Bonne, 2007). We therefore hypothesized that LMNA in normal embryos and proliferating cancer cells mechanosenses the microenvironment and increases not only to stiffen the nucleus (Osmanagic-Myers, et al., 2015; Dahl, et al., 2008; Pajerowski, et al., 2007) but also to regulate repair factors that confer resistance to both DNA damage and cell cycle arrest.

2.3 Results

2.3.1 Turnover of Lamin-A is affected by phosphorylation of the protein at specific sites

Phosphorylation of LMNA by CDK1 and PKC at specific sites to disassemble the lamina at the onset of mitosis is well established(Dechat *et al.*, 2010), however phosphorylation has also been observed during interphase. Transduction followed by clone isolation of point mutated LMNA A549 cell line was used to study the phenotype of phosphorylation on

LMNA. S22E LMNA is a phospho-mimetic LMNA with serine 22 site mutated to glutamic acid, whereas S22A mutant LMNA cannot be phosphorylated at serine 22 since alanine cannot be phosphorylated. **Fig 2.1a-i** shows phosphor-mimetic LMNA shows equal LMNA distribution between the nuclear periphery and the nucleoplasm whereas S22A is concentrated on the periphery indicating that phosphorylation during the interphase causes LMNA to dissociate from the nuclear lamina. To show that this is not due to differences in protein translation but post-translational modification followed by degradation, we treat the cells with translation inhibitor – cycloheximide (Trx-i) for 24 hours and observe that there is no difference after treatment, with both S22E samples have consistently higher nucleoplasm to lamina ratio (**Fig2.1a-ii,iii**). Moreover, anti-GFP and anti-LMNA immunoblots of S22E mutant with or without cycloheximide treatment stain for fragments at lower molecular weight which are not present in the S22A mutant, indicating further degradation of the phosphor-mimetic LMNA after being solubilized in the nucleoplasm (**Fig 2.1b**). In an interesting concurrence, we find that metalloproteases like MMP2 that cleave the extracellular matrix proteins like collagen-I, also degrade phosphorylated Lamin-A, as shown by the rescue of S22E LMNA levels by treating cells with MMP2-I (**Fig 2.1c**), exhibiting an interesting link between the extracellular matrix and lamin-A that has previous been studied as a scaling law across many different tissues with varying stiffness(Swift *et al.*, 2013). Immunofluorescence with anti-LMNA show increased nuclear LMNA in S22E mutant LMNA cells treated with MMP-i indicating LMNA rescue at the individual cell level (**Fig 2.1d**). Since the cytoskeleton links the ECM to the nuclear membrane through LINC complex, we study if acto-myosin stress also affects the phosphor-mimetic LMNA phenotype. We use myosin II inhibitor- Blebbistatin to inhibit acto-myosin stress which has previously been shown to round up cells cultured on plastic (**Fig 2.1e**). We then stain the cells with γH2AX and observe the number of foci (indicator of double stranded DNA break) (**Fig 2.1f**). The functional consequence of depleted Lamin-A at the nuclear lamina due to its phosphor-mimetic form can be seen by the increase in γH2AX foci observed in cells

plated on stiff plastic (**Fig 2.1g-i**). This accumulation of DNA damage is rescued in both S22E and S22A LMNA mutant by treatment with blebbistatin, indicating the increased DNA damage consequences of increased nucleoplasmic LMNA are downstream of acto-myosin stress. Moreover, this increase in DNA Damage results in cell cycle arrest as seen by the increased accumulation of S22E LMNA cells in G1 phase of cell cycle (**Fig 2.1g-ii**). The cell cycle arrest is again rescued by acto-myosin stress inhibition. We determine that cells in G0/G1 phase of cell cycle are sensitive to acto-myosin stress inhibition and round up exhibiting decreased nuclear area, this effect is rescued by washing out of the drug (Fig 1h). The phenotype of the S22E LMNA mutant can be compared to a cell line with knockdown of LMNA (shLMNA), with both cell lines showing increased DNA damage compared to the GFP WT LMNA (wild type/no mutation (WT) LMNA with GFP added at the beginning) and S22A LMNA, which is then rescued by acto-myosin inhibition. Given that acto-myosin stress inhibition can rescue effects of phosphor-mimetic LMNA, we further investigated if acto-myosin stress can itself affect LMNA levels. Moreover, what is the mechanistic cause of this accumulation of DNA damage and cell cycle arrest in phospho-mimetic LMNA?

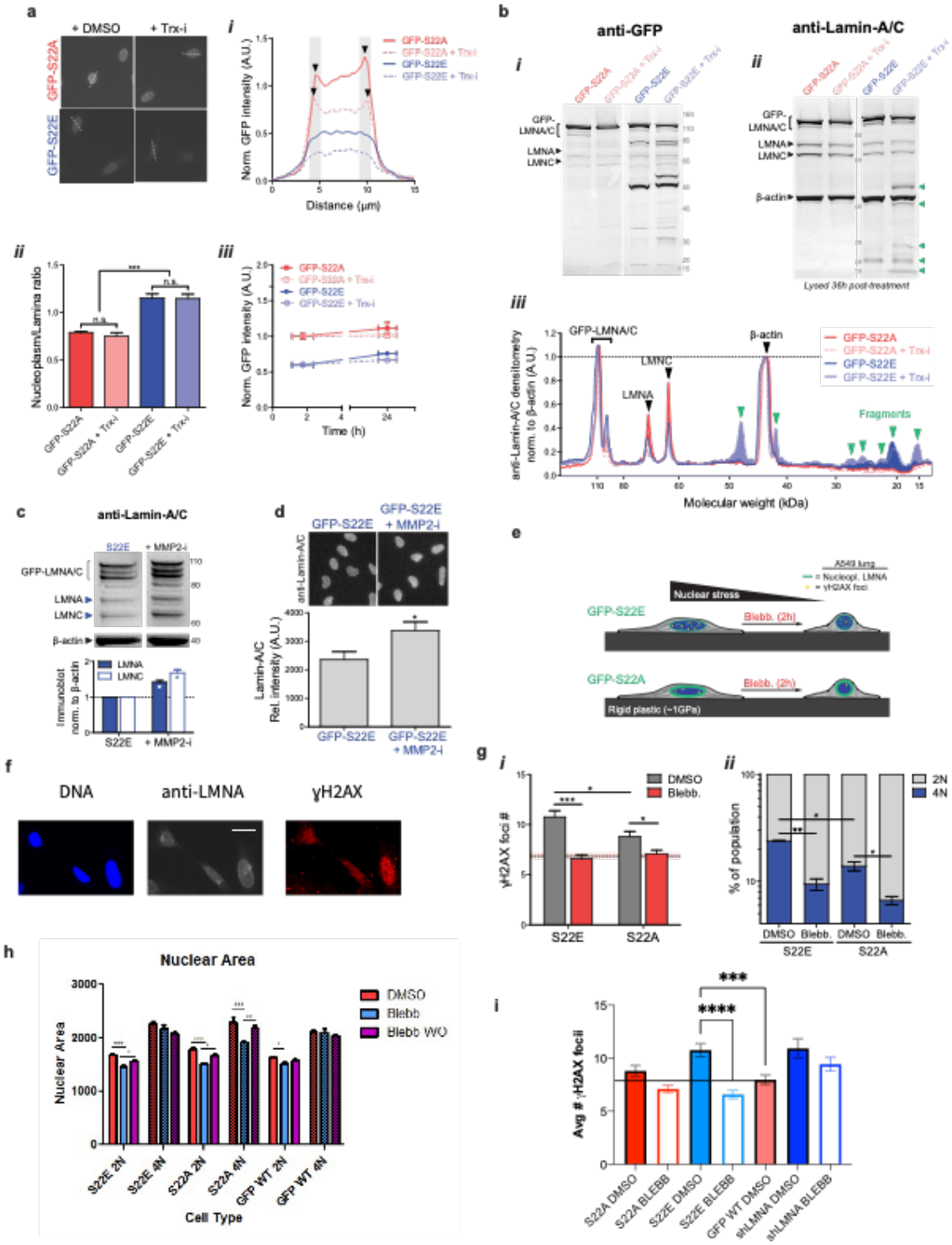


Figure 2-1 Phospho-mimetic LMNA cells show accumulated DNA damage and cell cycle arrest

a(*i*) Phospho-mimetic GFP-S22E LMNA shows higher intensity of GFP in the nucleoplasm compared to GFP-S22A LMNA mutant. **(*ii,iii*)** Inhibition of protein translation does not affect these differences indicating the cause to be post-translational modification **b)** Phospho-mimetic mutant GFP-S22E (± Trx-i, 3h) probed with **(*i*)**anti-LMNA and **(*ii*)** anti-GFP shows low LMNA versus GFP-S22A cells. Stable lines were made within a stable shLMNA line to minimize both endogenous LMNA and overexpression artifacts. Low-MW fragments (green arrows) are not detected in GFP-S22A cells. **c)** MMP2-i rescue S22E levels revealing MMP2 to play a role in degradation of LMNA **d)** Immunofluorescence with anti-LMNA confirms rescue of nuclear LMNA levels through MMP2-i at the individual cell level **e)** Schematic of A549 cells rounding up when treated with Blebbistatin **f)** Representative images of GFP LMNA A549 cells stained for anti-LMNA and anti-γH2AX (Scale bar=20μm) **g(*i*)** S22E LMNA cells show increased accumulated basal DNA damage compared to S22A cells which is then rescued by Blebbistatin and **(*ii*)** %4N cells are highest in S22E cells, but blebb minimizes differences **h)** S22E, S22A and WT LMNA in G0/G1 phase show more rounded nuclei with less area when treated with Blebbistatin which is then rescued by wash out of the drug **i)** Increased DNA damage in S22E cells is similar to that observed in cells with knockdown of LMNA which is then rescued by Blebbistatin (n>93 nuclei per cond.; t-test *p<0.05, **<0.01, ***<0.001; irrelevant blot lanes have been removed and is indicated by a blank space

2.3.2 Lung Adenocarcinomic cells show increased mislocalization of DNA repair factors after LMNA knockdown

To study how mechanical stress can cause accumulation of DNA damage in mutant Lamin-A cells, we hypothesize that one of the functions of Lamin-A is to maintain the integrity of the nuclear membrane, decreased Lamin-A at the lamina in combination with mechanical can increase the probability of nuclear membrane rupture, which might have adverse effects like mis-localization of nuclear proteins including DNA repair factors leading to accumulation of DNA damage. To test this hypothesis, we use a A549 cell-line with stable knockdown of LMNA - shLMNA A549 cells. We immunostain the cells with a nuclear protein – DNA repair factor which plays a key role in the NHEJ pathway- Ku80 (**Fig 2.2a**). We observe 5.8% of cells show Ku80 which is mis-localized to the cytoplasm. We also immunostain for Lamin-A to verify that this mis-localization is not due to mitosis and these cells are in interphase when the rupture occurs. We find that inhibiting the acto-myosin stress using Blebbistatin decreased the probability of observing cells with nuclear membrane rupture. Interestingly, however, when the Blebbistatin was washed away to restore acto-myosin stress, there was a significant spike in the number of ruptured nuclei observed. Moreover, the spike is significantly increased when the cells are treated with a nuclear import inhibitor-Ivermectin (Imp-i). This increase in % of nuclei with mis-localized Ku80 is not observed when the cells are

treated with nuclear import inhibitor mifepristone which is specific to the Integrase nuclear import (HIV) but not Ku80. These studies are further elaborated in Chapter 4.

2.3.3 Lamin-A levels modulated based on acto-myosin stress effects nuclear membrane rupture probability and subsequently DNA Damage and cell cycle arrest in ex-vivo embryonic hearts

To study if Lamin-A and acto-myosin stress dependent nuclear membrane rupture is also observed in *ex vivo* somatic proliferating cells, we study embryonic chicken hearts which can independently beat ex-vivo for 4-5 days. To detect nuclear membrane rupture in the hearts, we transfect the hearts with GFP-Ku80 as well as mCherry-cGAS. cGAS (cyclic GMA-AMP synthase) is a cytosolic protein that binds DNA. It is part of the cGAS-STING pathway, which evolved as cellular immune system that triggers an inflammatory response (type I interferon response), by binding to viral DNA, intracellular bacteria DNA or even the DNA from the cell's own nucleus. We use it to detect nuclear membrane rupture in ex-vivo hearts by identifying cells that are positive for Ku80 in the cytoplasm and puncta of c-GAS at the nuclear periphery. Confocal images show that hearts treated with Blebbistatin for 2 hours show significant reduction in the % of cells with nuclear membrane rupture (**Fig 2.2b**). Interestingly, immunostaining of Lamin A in those hearts also shows a significant drop, indicating that Lamin A is "mechanosensitive" and is acutely responsive to reduction in acto-myosin stress. We also find that sarcomeres in the heart are disrupted by myosin II inhibition, confirmed by staining of actinin α2, though they rapidly recover after 1 hour from washing out Blebbistatin. The hearts start beating again within minutes of washout of drug. However, Lamin-A levels do not recover after 1 hour of washout, presumably because reassembly of Lamin-A takes a longer time (it has previously been reported to be 90-180 minutes(Goldman *et al.*, 1992)) or Lamin-A gets degraded in response to lack of acto-myosin stress. Interestingly, in this situation of acute reintroduction of acto-myosin stress with reduced Lamin-A at the lamina, there is a significant spike in the probability of observed nuclear membrane rupture. Moreover, when the hearts are treated with Imp-i after washing Blebbistatin out, we find that there is sustained increase in probability of observing mis-localized Ku80 after 24 hours (**Fig 2.2c-*i***), though the drug washout hearts return to

control levels of nuclear membrane rupture. Embryonic hearts where LMNA has been knockdown using vivo-Morpholinos also show significantly increased mis-localized Ku80 probability which is rescued by inhibition of acto-myosin stress. Acute as well as prolonged mis-localization of DNA repair factors like Ku80 due to the rupture causes increase in double stranded DNA damage, indicated by increase γH2AX (**Fig 2.2c-*ii***). The treatment with the two drugs does not significantly increase the number of apoptotic cells with fragmented DNA (**Fig 2.2d**). More importantly, this increase in DNA Damage results in cell cycle arrest in these beating hearts after blebbistatin treatment. Percentage of cells in different phases of the cell cycle can be measured by introducing EdU in the media one hour before fixing the heart tissue, cells in S phase incorporate EdU (which can be fluorescently labelled) since it is the analog of DNA base pair thymine (T). EdU positive cells can show puncta indicating sites of active DNA replication (**Fig 2.2e**). Using EdU intensity as well as DNA intensity, one can identify if the cell has 2N or 4N chromosomes, hence the cells can be identified as being in G1, S or G2 phase (**Fig 2.2f**). The blebbistatin washout hearts show significant accumulation of cells in the G2 phase indicating cell cycle arrest (**Fig 2.2g**). Hence, somatic tissue cells, unlike some tumor cells, are sensitive to accumulation of DNA damage in presence of impaired Lamin-A, resulting in cell cycle arrest.

Reversibility of acto-myosin inhibition is illustrated not only in the sarcomeres of the heart, but also cell nuclei. In the absence of acto-myosin stress, the nuclei round up, exhibiting significantly increased circularity, however 1 hour after washing out blebbistatin, the cells return the prior circularity of beating heart (**Fig 2.2h**). This indicates that circularity of nuclei may depend on acto-myosin stress more than Lamin-A levels since Lamin-A is still low 1 hour after Blebbistatin wash out.

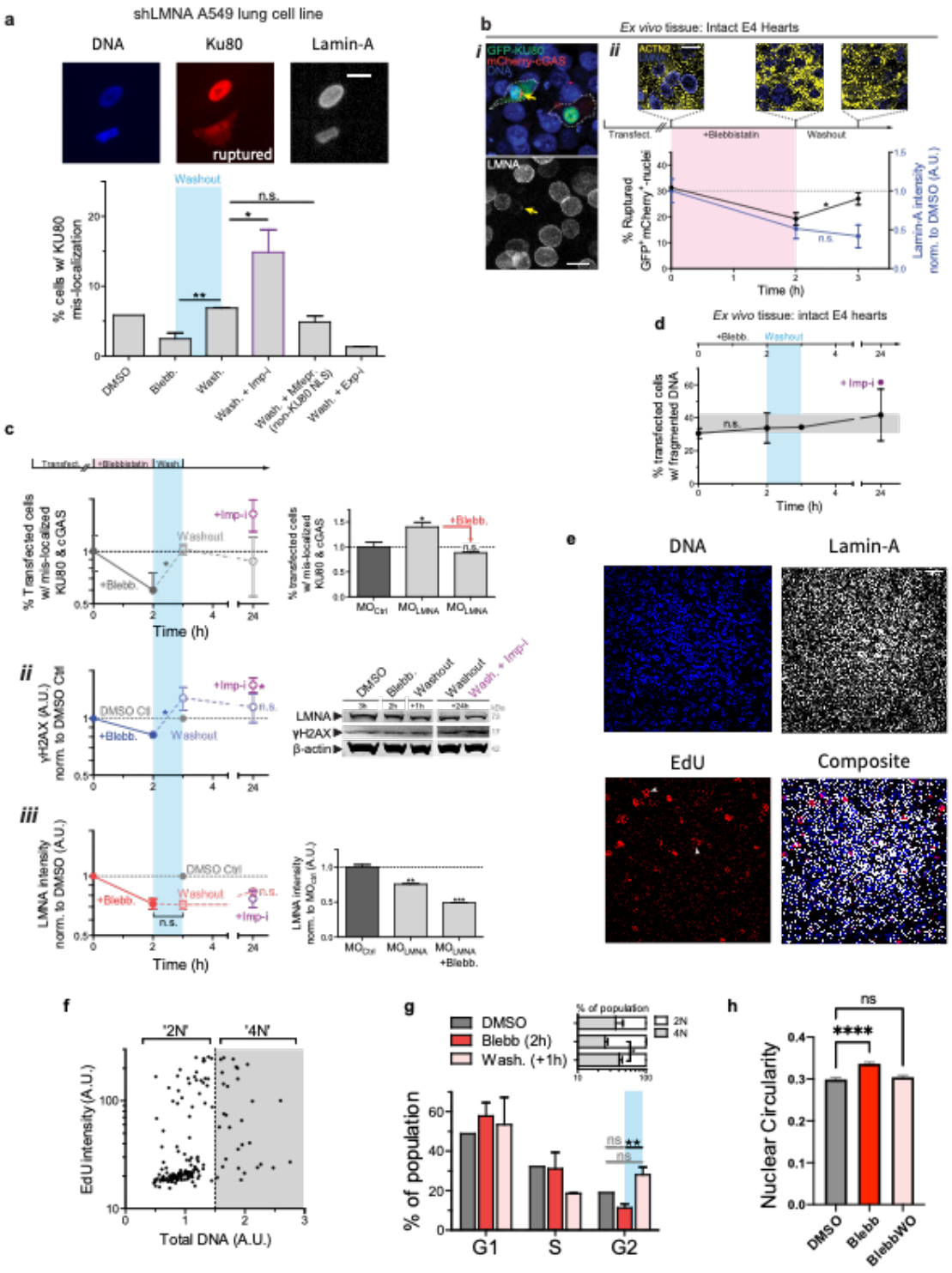


Figure 2-2 'Mechanosensitive' Lamin-A protects from mis-localization of DNA repair factors and subsequent DNA damage and cell cycle arrest

a) Representative image of mis-localized Ku80 in an interphase cell (Scale bar = 20μm). Blebbistatin washout results in an increase in rupture with cytoplasmic mis-localization of KU80. An excess of rupture is seen with washout plus a nuclear import inhibitor (ivermectin, 'Imp-i', which inhibits Impα/β mediated import), but not with washout plus a specific inhibitor of integrase (IN) nuclear import (mifepristone, 'Mifepr.') that does not target KU80's NLS. (t-test *p<0.05, **<0.01). All error bars indicate SEM. **b(i)** Images of E4 hearts transfected with GFP-KU80 and mCherry-cGAS. Arrow: cell with low LMNA and ruptured nucleus, with cytoplasmic mis-localization of GFP-KU80 and mCherry-cGAS puncta at nuclear envelope. **(ii)** Images of striation (α-actinin-2) before blebb, 2h after, and upon washout (1h). Blebbistatin treatment results in reversible inhibition of contractility. %-Ruptured nuclei (double +'s) decreases with myosin-II inhibition but increases upon washout. **c(i)** LMNA KD increases mis-localized KU80 and cGAS but is rescued by myosin-II inhibition. LMNA immunofluorescence with MO_{LMNA} ± blebb. (n>139 cells per cond.). Mis-localized KU80 increases further with import inhibitor ivermectin, Imp-i. (n>111 double-transfected cells per cond'n.; t-test: *p<0.05)(ii) anti-γH2AX and **(iii)** anti-LMNA immunoblot of blebb-treated E4 hearts after washout ± Imp-i (8 hearts per lysate; t-test: *p<0.05). **d)** Blebbistatin treatment and washout have no significant effect on cell death/viability, as determined by %-transfected cells with fragmented DNA. **e)** Representative z-stack of confocal images of embryonic hearts showing EdU positive cells **f)** Representative plot of EdU vs DNA intensity for cell cycle phase (G1 vs S vs G2; or '2N' vs '4N'). **g)** Blebb washout increases G2 and 4N cells (vs 2N). n>77 cells per cond.; t-test: **p<0.01. **h)** Cardiomyocytes in Blebb treated hearts are significantly more circular (****p<0.0001, n>638 cells analyzed per sample)

2.3.4 Increase in DNA damage due to knockdown of Lamin A emulates knockdown of DNA repair factors

Adding an external DNA damage source like etoposide which stabilizes topoisomerase IIA (a protein that nicks DNA during damage repair and transcription) and hence increases DNA damage allows us to establish the timeline of DNA repair. We find that U2OS cells which have transfected with siLMNA show similar impairment in DNA damage repair as cells where a combination of DNA repair factors have been knocked down- siKu80, siBRCA1, siBRCA2 or even just siKu80. This impairment persists even in the presence of an external DNA damage agent and returns to the steady state levels five hours after the induction of increased DNA damage (**Fig 2.3a**). Increased DNA damage can have functional consequences to the phenotype of the beating heart. We use human induced pluripotent stem (hiPS) cells and differentiate them into cardiomyocytes, which form organoids and 'beat' in the dish. Kymographs show that organoids treated with etoposide for 2.5hrs show arrhythmia, moreover, after washing out the DNA damage agent, the beating does not regain original beats per minute (bpm) even after 24 hours (**Fig 2.3b**).

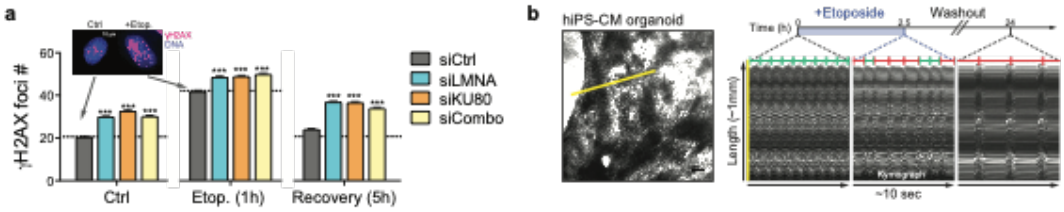


Figure 2-3 External DNA Damage agents impair beating heart phenotype

a) KDs increase γH2AX before/during/after 10 μM etoposide (n>39 cells per cond). **b)** Kymograph of beating hiPS-CM organoids generated by tracing length along yellow line with time.

2.3.5 Lamin A plays a role in telomere maintenance

Non-phosphorylated(S22A) Lamin A (**Fig 2.4a**) and phosphorylated(S22E) Lamin B1 (**Fig 2.4b**) form structures that resemble puncta in the nucleoplasm which have been reported to be higher in number during G1/S phase of the cell cycle (Goldman *et al.*, 1992). We posit the hypothesis that it is plausible for phosphorylated(solubilized) and non-phosphorylated(non-solubilized) form of Lamin A to directly associate with nuclear proteins like DNA repair proteins, even away from the membrane lamina.

Given our studies on mis-localization of Ku80 with Lamin-A knockdown (**Fig 2.4c**), we decided to study if the Ku protein complex and other DNA repair proteins directly interact with Lamin-A, and if so, which form of Lamin-A (phosphorylated or non-phosphorylated) do they preferentially interact with. We conducted coimmunoprecipitation (Co-IP) experiments by pulling down GFP antibody and pS22 LMNA antibody on GFP WT LMNA, GFP S22E LMNA, GFP S22A LMNA, WT LMNA A549 cell lines followed by mass spectrometry to analyze the proteins in the supernatant and pull-down. We used rabbit monoclonal antibodies, so we used rabbit immunoglobin (IgG) as the control antibody for pull-down. This is to check that any pull-down from the antibody is due to the specific interaction between the antibody and the precipitated protein, and not due to non-specific interactions with the rabbit protein. 71.5% of the proteins detected in the pull-down from GFP antibody are different from the pull-down from IgG, confirming that GFP pull down came from

unique interactions (**Fig 2.4d**). We found that 91.2% of the proteins detected in the supernatant of the pull down from GFP and IgG are common, again confirming that only GFP antibody specific interactions were pulled down. We found that while some DNA repair proteins like Ku70, RPA1 interacted with both forms of LMNA, Ku80 preferentially interacted with only the non-phosphorylated form of LMNA which is present in the nuclear lamina since Ku80 did not precipitate with the pS22 pull down of WT LMNA, neither did it precipitate with the GFP pull down of the GFP S22E cell line (**Fig 2.4e**).

Hutchinson–Gilford progeria syndrome (HGPS) is a rare genetic disease caused by a mutation in LMNA gene which has symptoms of accelerated aging in which most patients suffer from cardiovascular problems. In an interesting synchronicity, Ku80 deficient mice show progeroid symptoms(Reiling *et al.*, 2014). Moreover, Ku80 has been shown to be involved in maintaining telomere length as well as recruitment of an important shelterin complex protein TRF2(Boulton and Jackson, 1998; Melnikova, Biessmann and Georgiev, 2005; Fink *et al.*, 2010). This prompted us to study the effect of mis-localization of Ku80 due to nuclear membrane rupture on telomere length (**Fig 2.4f**). We used Dox inducible gene edited mCherry TRF1 U2OS cells and measured intensity of TRF1 foci in ruptured and non-ruptured cells by developing our own algorithm. We observe that ruptured nuclei show greater telomere attrition which is similar to the effect seen in Ku80 knockdown cells (**Fig 2.4g,h**). Moreover, when we do Telomere Q-FISH staining on cardiomyocytes which have knockdown of LMNA using morpholinos, we observe telomere attrition which is rescued by treatment with blebbistatin (**Fig 2.4i**). Hence LMNA a directly interacts with Ku80 and knockdown of LMNA results in telomere attrition which can contribute to accelerated aging.

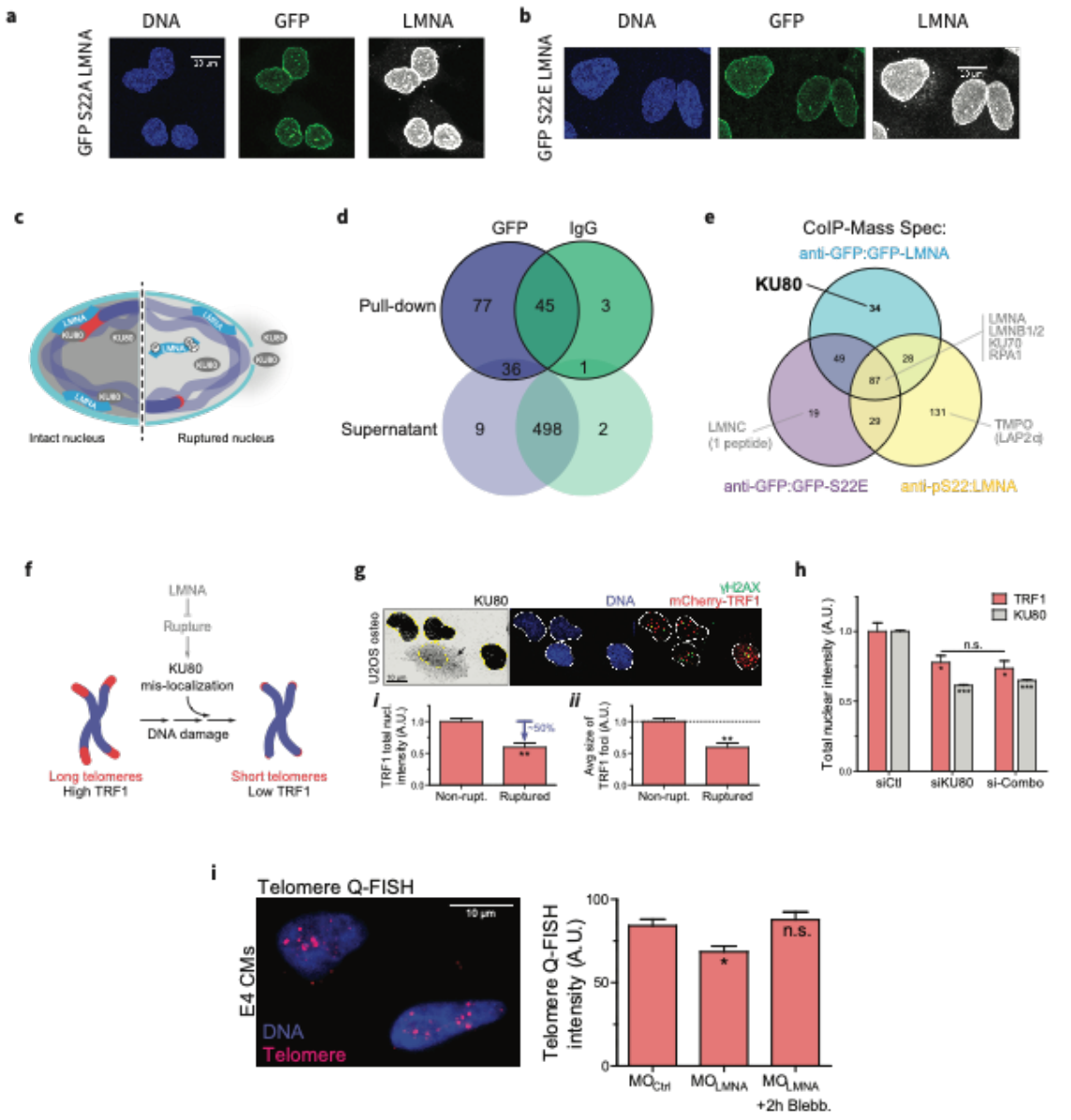


Figure 2-4 Lamin-A knockdown cause telomere attrition

Representative confocal images of **a)** GFP-S22A Lamin-A and **b)** GFP-S22E Lamin A showing foci in the nucleoplasm region **c)** Schematic: effect of repair factor mis-localization on telomeres **d)** Venn Diagram showing overlap of precipitated and supernatant proteins in GFP and IgG **e)** CoIP-MS of WT LMNA, phosphorylated LMNA (anti-S22), and phospho-mimetic LMNA (GFP-S22E). **f)** Cartoon: LMNA-KU80-telomere interactions at the lamina **g)** U2OS expressing mCherry-TRF1. Ruptured nuclei (n=7) with cytoplasmic KU80 exhibit (i) lower TRF1 and (ii) smaller TRF1 foci (t-test: *p<0.05). **h)** siKU80 and si-Combo decrease TRF1 (n>89 cells per cond.). **i)** Telomere Q-FISH in embryonic CMs shows MO_{LMNA} shortens telomere unless rescued by blebb (n>81 nuclei per cond.). (Unless noted, one-way ANOVA: *p<0.05, ***<0.001)

2.4 Discussion

LMNA thus mechano-protects by retaining DNA repair factors in the nucleus (**Fig 2.2, 2.4**), which thereby prevents excessive accumulation of DNA damage in stiff microenvironments and/or under conditions of high actomyosin contractility. Rapid change in LMNA protein independent of any transcription/translation (Trx-i) rules out many possible contributing pathways to the equally rapid DNA damage response. Such conclusions about causality are difficult to otherwise achieve with experiments conducted over many hrs/days such as with mouse mutants. Lack of protection by LMNA and consequent loss of DNA repair factors could compromise genome integrity when actomyosin stress is high (**Fig. 2.2, 2.4**), but a source of DNA damage might also increase upon rupture. Enhanced entry of cytoplasmic nucleases(Maciejowski *et al.*, 2015) or other cytoplasmic factors (e.g. reactive oxygen species) cannot be ruled out. However, the γH2AX foci are typically distributed throughout the nucleoplasm (**Fig 2.1f, 2.3a, 2.4g**) – whereas localization of cytoplasmic proteins that bind DNA strongly (e.g. cGAS) are restricted to the rupture site. Rupture consistently results in 'global' mis-localization of multiple DNA repair factors in several pathways(Irianto *et al.*, 2017; Xia, Ivanovska, *et al.*, 2018), and upon repair factor depletion, DNA damage is in excess before, during, and after a transient increase in damage (by 1h etoposide, **Fig.2.3a**).

LMNA defects cause disease through "cell-extrinsic mechanisms" that likely include ECM and/or cytoskeletal stress. Mosaic mice in which 50% of the cells express defective LMNA maintain a normal lifespan, whereas mice with 100% defective cells die within weeks of birth(de la Rosa *et al.*, 2013). Cultures on rigid plastic of the same cells (and similar cell types(Hernandez *et al.*, 2010)) exhibit premature senescence/apoptosis, as is common with excess DNA damage, but growth and viability are surprisingly rescued upon culture on almost any type of ECM. Reduced cytoskeletal stress and suppresses nuclear rupture and DNA damage in embryonic CMs and immortalized cancer lines with low LMNA (**Fig 2.2a,b**). The findings are further consistent with the observation that laminopathies spare soft tissues such as brain, independent of lineage or developmental origin, but generally affect stiff and mechanically stressed adult tissues including muscle or bone(Cho, Irianto and Discher, 2017).

2.5 Materials and Methods

2.5.1 Embryonic chick hearts and cardiomyocytes

White Leghorn chicken eggs (Charles River Laboratories; Specific Pathogen Free (SPF) Fertilized eggs, Premium #10100326) were used to extract embryonic hearts. SPF chicken eggs from Charles River are produced using filtered-air positive-pressure (FAPP) poultry housing and careful selection of layer flocks. Every flock's SPF status is validated in compliance with USDA memorandum 800.65 and European Pharmacopoeia 5.2.2 guidelines. SPF eggs were incubated at 37˚C with 5% CO_2 and rotated once per day until the desired developmental stage (e.g. four days for E4; Hamburger-Hamilton stage 23-24 (HH23-24)). Embryos were extracted at room temperature (RT) by windowing eggs, carefully removing extra-embryonic membranes with sterile sonicated forceps, and cutting major blood vessels to the embryonic disc tissue to free the embryo. The extracted embryo was then placed in a dish containing PBS on a 37˚C-heated plate, and quickly decapitated. For early E2-E5 embryos, whole heart tubes were extracted by severing the conotruncus and sino venosus. For more mature (>E5) embryos, embryonic discs were extracted by windowing the egg, cutting out the embryo with the overlying vitelline membrane intact, lifting out the embryo adherent to the vitelline membrane and placing in a dish of pre-warmed PBS. Extra-embryonic tissue was carefully cut away using dissection scissors and the embryo was teased away from the vitelline membrane using forceps. Whole hearts (>E5) were extracted by severing the aortic and pulmonary vessels. The pericardium was carefully sliced and teased away from the ventricle using extra-fine forceps. E10 brain and liver tissue were collected from the presumptive midbrain and hepatic diverticulum, respectively. All tissues were incubated at 37˚C in pre-warmed chick heart media (α-MEM supplemented with 10 % FBS and 1% penn-strep, Gibco, #12571) for at least 1 h for stabilization, until ready for use.

2.5.2 Induced pluripotent stem cell-derived cardiomyocytes (iPS-CMs)

Normal human iPS cells (obtained from Dr. Joseph Wu, Stanford Cardiovascular Institute (CVI) BioBank) were cultured following the protocol provided: Matrigel (BD Matrigel, hESC qualified: #354277) was suspended in cold 4°C DMEM/F12 medium 1:200 dilution (DMEM/F12 medium #10-092-CM-Fisher), mixed gently, and 1ml of this suspension was added to one 6-well plate (Corning Catalog #353046) and incubated for 1hr at RT to allow Matrigel to coat the surface. The solution was gently aspirated and small aggregates of human iPS cells were added to each well in mTesr1 medium containing 10 μM of ROCK inhibitor (Y27632, 2HCl – 50 mg: #50-863-7-Fisher). Culture medium was replaced daily (no ROCK inhibitor) until the cells reached ~85% confluency.

Human induced pluripotent stem cells (hiPSCs) were differentiated into cardiomyocytes (hiPS-CMs) using the "Cardiomyocytes differentiation and maintenance kit" from Stem Cell technologies (#05010 & #05020). The differentiation process was followed as described by the manufacturer's protocol. Briefly, the mTesr1 medium with ROCK inhibitor (10 μM) was replaced with differentiation medium A, cultured for 2 days (Day 0), then subsequently to medium B for 2 days (Day 2) and again switched to medium C twice (Day 4 & 6). From Day 8 onwards, maintenance medium (Day 8) was added/refreshed every 2 days until spontaneous CM beating was observed through imaging.

2.5.3 Cell lines

A549 (human lung adenocarcinoma) cell lines were obtained from ATCC (American Type Culture Collection, Manassas, VA, USA) and U2OS cell lines were obtained from the laboratory of Roger Greenberg, University of Pennsylvania. ATCC provides cell line authentication test recommendations per Tech Bulletin number 8 (TB-0111-00-02; 2010). These include: microscopy-

based cell morphology check, growth curve analysis, and mycoplasma detection (by DNA staining) which were conducted on all cell lines used in these studies. All cell lines maintained the expected morphology and standard growth rates with no mycoplasma detected. U2OS and A549 cell lines were cultured in DMEM high-glucose media and Ham's F12 nutrient mixture (Gibco, Life Technologies), respectively, supplemented with 10% fetal bovine serum (FBS) and 1% penicillin and streptomycin (Sigma-Aldrich).

2.5.4 Mass spectrometry (LC-MS/MS) of whole heart lysates

Mass spectrometry (LC-MS/MS, or 'MS') samples were prepared using procedures outlined in Swift et al.. Briefly, ~1 mm^3 gel sections were carefully excised from SDS–PAGE gels and were washed in 50% 0.2 M ammonium bicarbonate (AB), 50% acetonitrile (ACN) solution for 30 min at 37°C. The washed slices were lyophilized for >15 min, incubated with a reducing agent (20 mM TCEP in 25 mM AB solution), and alkylated (40 mM iodoacetamide (IAM) in 25 mM AB solution). The gel sections were lyophilized again before in-gel trypsinization (20 mg/mL sequencing grade modified trypsin, Promega) overnight at 37°C with gentle shaking. The resulting tryptic peptides were extracted by adding 50% digest dilution buffer (60 mM AB solution with 3% formic acid) and injected into a high-pressure liquid chromatography (HPLC) system coupled to a hybrid LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) via a nano-electrospray ion source.

Raw data from each MS sample was processed using MaxQuant (version 1.5.3.8, Max Planck Institute of Biochemistry). MaxQuant's built-in Label-Free Quantification (LFQ) algorithm was employed with full tryptic digestion and up to 2 missed cleavage sites. Peptides were searched against a FASTA database compiled from UniRef100 gallus gallus (chicken; downloaded from UniProt), plus contaminants and a reverse decoy database. The software's decoy search mode was set as 'revert' and a MS/MS tolerance limit of 20 ppm was used, along with a false discovery rate (FDR) of 1%. The minimum number of amino acid residues per tryptic peptide was set to 7, and MaxQuant's 'match between runs' feature was used for transfer of peak identifications across samples. All other parameters were run under default settings. The output tables from MaxQuant were fed into its bioinformatics suite, Perseus (version 1.5.2.4), for protein annotation and sorting.

2.5.5 siRNA knockdown and GFP-repair factor rescue

All siRNAs used in this study were purchased from Dharmacon (ON-TARGETplus SMARTpool; siBRCA1, L-003461-00; siBRCA2, L-003462-00; siKu80, L-010491-00; siRPA1, L-015749-01; siLMNA, L-004978-00 and non-targeting siRNA, D-001810-10). GFP-KU70 and GFP-KU80 were gifts from Dr. Stuart L Rulten from University of Sussex, Brighton, UK, and GFP-53BP1 was a gift from Dr. Roger Greenberg from University of Pennsylvania. Cells were plated 24 hours prior to transfection. Lipofectamine/nucleic acid complexes were prepared according to the manufacturer's instructions (Lipofectamine 2000, Invitrogen), by mixing siRNA (25 nM) or GFPs (0.2-0.5ng/ml) with 1 µg/ml Lipofectamine 2000. Final solutions were added to cells and incubated for 3 days (for siRNAs) or 24 hours (for GFPs) in corresponding media containing 10% FBS.

2.5.6 mCherry D450A FOK1 TRF1 expression

Death domain (DD)– Oestrogen receptor (ER)–mCherry–TRF1–FokI constructs were cloned and concentrated TRF1-FokI lentivirus with polybrene (8 mg/ml) diluted in media was added to U2OS cells at a minimum titer. Doxycycline-inducible TRF1–FokI lines were generated using the Tet-On 3G system. Doxycycline was used at a concentration of 40 ng/ml for 16–24 h to induce expression of TRF1– D450A FokI. Shield-1 (Cheminpharma LLC) and 4-hydroxytamoxifen (4-OHT) (Sigma) were both used at a concentration of 1 µM for 2 h, to allow for TRF1–D450A FokI stabilization and translocation into the nucleus.

2.5.7 Quantitative Fluorescence *In Situ* Hybridization and Telomere Analysis

Isolated cells were attached on slides, fixed and stained for PNA telomeric probe as we previously described in(Tichy *et al.*, 2017). Images of mMuSCs were taken using a Nikon eclipse 90*i* wide-field epifluorescence microscope equipped with a Prior Proscan III motorized stage, a Photometrics Coolsnap HQ2 14-bit digital camera, and a Nikon 100×/1.40 Plan Apo VC objective. Telomeres were analyzed with the investigators blinded to genotypes and/or conditions using open-source software Telometer, as previously described. The program generates statistics on the entire region of the nucleus. Analysis includes the intensity sum of all Cy5 telomere pixels for a given nucleus (proportional to the cell's total telomere length) and the intensity sum of all DAPI pixels for the nucleus (proportional to total cellular nuclear DNA content).

2.5.8 Confocal Imaging

For confocal imaging of embryonic heart tissue, heart samples were rinsed with PBS, fixed with 4% paraformaldehyde (PFA, Fisher) for 45 min, washed x3 with PBS, and permeabilized with 0.5% Triton-X (Fisher) in PBS for 3 hrs. Samples were then incubated overnight with primary antibodies in 0.5% BSA solution with gentle agitation at 4 °C (as described above), and washed x3 in 0.1% BSA and 0.05% Triton-X in PBS. Corresponding secondary antibodies were added at 1:500 dilution for 1.5 hrs at RT. Immunostained hearts were then mounted between glass coverslips with mounting media (Invitrogen ProLong Gold Antifade Reagent). Images were taken with Leica TCS SP8 system with either a 63×/1.4 NA oil-immersion or 40×/1.2 NA water-immersion objective.

Chapter 3 Scaling concepts in 'omics: Lamin B1 is regulated by FOXM1 and predicts poor prognosis, unlike fibrosis

This chapter has been adapted from a paper submitted and in revision to PNAS.

Sangkyun Cho performed the mass spectrometry analysis on the 3 Liver cancer patient samples obtained with the help of Prof Rebecca Wells in Fig 3.2c,d. Jerome Irianto and Yuntao Xia performed the scRNA seq experiments on A549 cells in Fig 6. Brandon Hayes developed the CRISPR edited GFP-H2B A549 cell line and Mai Wang conducted the analysis in Fig 3.5a,b. Farshid Jafarpour in conjunction with Prof Andrea Liu wrote the supplementary theory section on Population scaling from single cell scaling

3.1 Abstract

Physicochemical principles such as stoichiometry and fractal assembly can give rise to characteristic scaling between components that potentially include co-expressed transcripts. For key structural factors within the nucleus and extracellular matrix (ECM), we discover specific gene-gene scaling exponents across many of the 32 tumor types in The Cancer Genome Atlas (TCGA), and we demonstrate utility in predicting patient survival as well as Scaling-informed Machine Learning (SIML). All cancers show tumor-elevated proliferation genes, and some subsets show scaling with the nuclear filament *LMNB1*, including the transcription factor FOXM1 that we verify in cancer lines. SIML shows such *regulated* cancers cluster together with longer overall survival than dysregulated cancers, but high *LMNB1* in half of regulated cancers surprisingly predicts poor survival, including for liver cancer. *COL1A1* is also studied because it too increases in tumors; and

a pan-cancer set of fibrosis genes shows sub-stoichiometric scaling with *COL1A1* but predicts patient outcome *only* for liver cancer -- unexpectedly predicting *longer* survival. Single-cell RNA-seq data shows non-trivial scaling consistent with power laws from bulk RNA and protein analyses, with SIML segregating synthetic from contractile cancer fibroblasts. Our scaling approach thus yields fundamentals-based power laws relatable to survival, function, and experiments.

3.2 Introduction

Dysmorphic nuclei prevail in tumors and sometimes reflect changes in nuclear lamins that influence nuclear shape and stiffness (Dahl *et al.*, 2005; Pajerowski *et al.*, 2007; Harada *et al.*, 2014a; Irianto, Charlotte R Pfeifer, *et al.*, 2016). Tumors also generally display abnormal architecture and often differ palpably from normal tissues. Breast tumor stiffness, for example, can be felt in self-exams and can affect tumor growth in 3D models (Levental *et al.*, 2009) (Meng *et al.*, 2018), and fibrotic rigidity is a major risk factor for liver cancer (Pinter *et al.*, 2016). Such physical changes are often attributed to fibrillar collagen accumulation in extracellular matrix (ECM), especially stoichiometric assembly of COL1A1 and COL1A2 proteins (Swift *et al.*, 2013; Buxboim *et al.*, 2014). In normal tissues, intriguingly, lamin-A/C (*LMNA*) scales with a characteristic power law versus collagen-I levels from soft brain to rigid bone, whereas Lamin-B1 (*LMNB1*) shows little variation normally but is elevated in some cancers including liver cancer(Sun *et al.*, 2010) (Abdelghany *et al.*, 2018). Here, in an analysis across all 32 tumor types in The Cancer Genome Atlas (TCGA) as well as the subset having both tumor and normal adjacent tissue data, we seek and indeed discover sets of genes that scale with the above nuclear and ECM factors, and we relate gene sets to patient survival and introduce scaling into machine learning.

Scaling approaches are reasonable to pursue because lamin and collagen polymers assemble into fractals with properties including stiffness that lead to *characteristic power law exponents* versus concentration (Swift *et al.*, 2013)(Yang, Leone and Kaufman, 2009) (e.g. **Fig.3.1a**). Correlative studies merely focus on ‘goodness of fit’ – whereas we seek specific scaling exponents that are reproducible and ‘universal’. Scaling pervades physics and includes, for example, fiber bending stiffness that scales with fiber thickness (~h^3), and scaling in biology

includes Klieber's power law(KLEIBER, 1947) (Thommen *et al.*, 2019) for metabolic rate ~ (body mass)$^{0.75}$. Scaling in chemistry can reflect stoichiometry (**Fig.3.1b**) and scaling might even apply to complex kinetics such as cell cycle (**Fig.3.1c**).

The *characteristic, non-trivial scaling* that we discover between some genes in TCGA is supported by our further experiments on tumor proteomics, diverse cancer cell line studies, and single-cell RNA-seq. We show, for example, with two major fibrillar collagen genes that *COL5A1* scales sub-stoichiometrically with *COL1A1*$^{\alpha}$ *across all 32 cancers* in TCGA with α = 0.84 (± 0.02, SEM) proving similar to bulk proteomics (α = 0.87) and single-cell RNA-seq (α = 0.88), thus illustrating a surprising universality across RNA, protein, bulk, and single cell measurements. We also present a novel means of Scaling-informed Machine Learning (SIML) that uses the small gene sets with characteristic scaling to better understand big data in TCGA (9112 patients) and single-cell RNA-seq.

3.3 Results

3.3.1 Survival of cancer cohorts segregate with Scaling-Informed Machine Learning (SIML)

TCGA's 9112 patients across 32 cohorts of cancer types (**Table 3.1**) have primary tumor mRNA expression data for 20530 genes, generating a large matrix that is well-suited to dimensionality reduction and visualization by Uniform Manifold Approximation and Projection (UMAP) (McInnes, Healy and Melville, 2018). The machine-learned clusters of patient cohorts distribute based on organ systems (**Fig.3.1d-i**); all tumors of the digestive system, for example, appear as nearest neighbors. Survival is not an input and no pattern of overall survival of patients with a given tumor type is evident; colon cancer has very limited median survival, for example, compared to the nearby cohort of stomach cancer. Moreover, adding tumor-adjacent uninvolved tissue data for the 16 cancers with matched data (**Table 3.1**), we find that the adjacent tissue clusters close to the corresponding primary tumor tissues (**Fig.S3.1a**).

Scaling-informed machine learning (SIML) was performed by repeating the above UMAP analysis with a minimal set of genes that we show below scale with *LMNB1* and that proves to be

cell cycle centric and includes one cell cycle transcription factor, *FOXM1*. Proliferation is of course a general hallmark of cancer, but such scaling will prove evident only in 19 tumor cohorts (of the 32), which we denote as 'regulated' (**Fig.3.1d-ii, Table 3.1**). Compared to the remaining 13 dysregulated tumor types, the regulated cancers show significantly longer median survival (**Fig.3.1d-iii**). SIML also better distinguishes adjacent univolved tissue for the 16 cancers with such data (**Fig.S3.1b**). This initial analysis motivates a stepwise elaboration of our scaling approach, with a further goal of assessing the relevance of gene-gene scaling to survival of a patient with a specific cancer type.

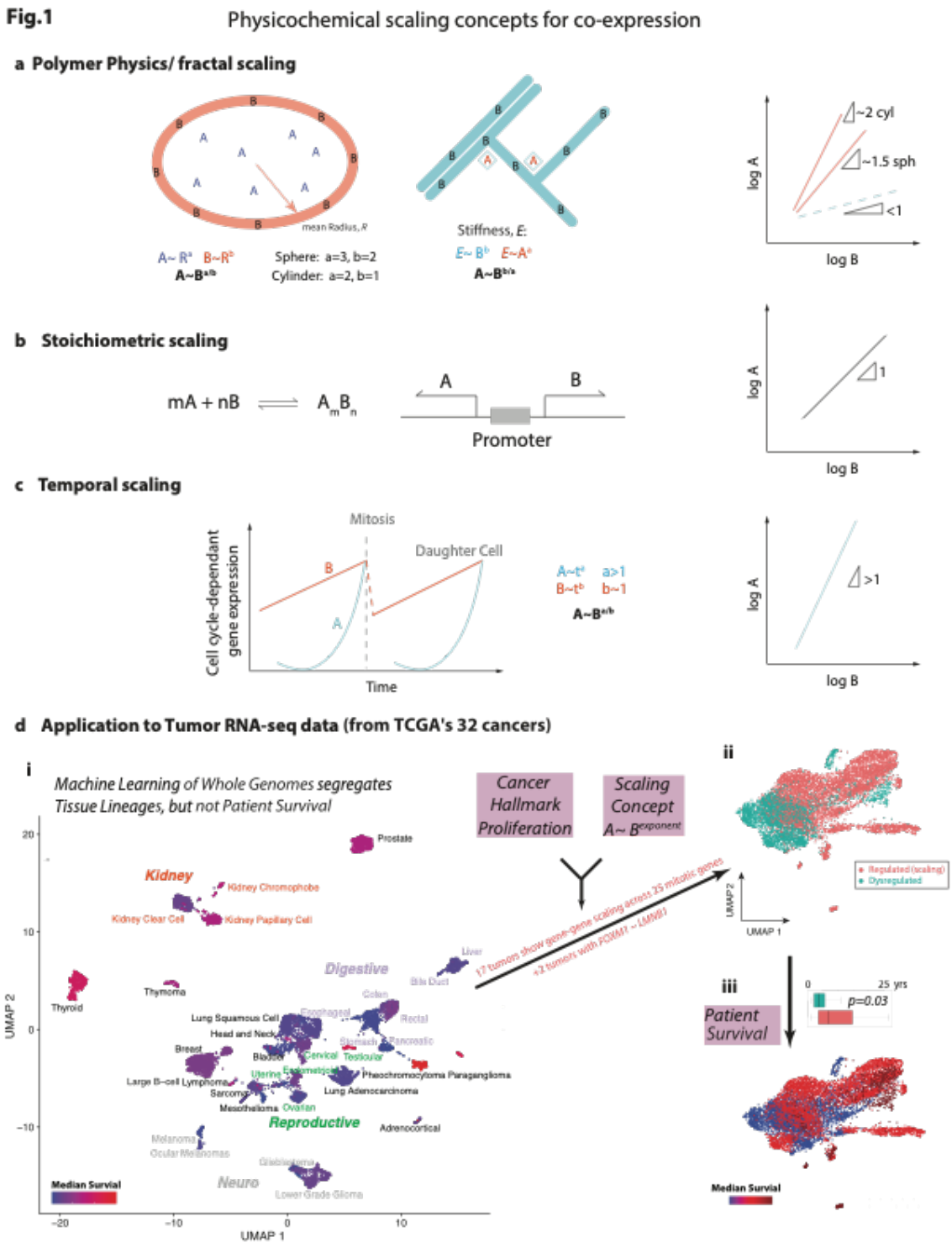


Figure 3-1 Physicochemical scaling concepts in gene-gene expression

a) Factors A and B distribute differently and scale per fractal physics of a volume and its surrounding surface or according to branched, decorated networks. Mechanical properties that

scale with such factors can also yield new scaling relations. **b)** Stoichiometric scaling is expected for co-regulated factors that form complexes or share a promoter. **c)** Genes that follow power laws in time (defined by phase of cell cycle) can also scale with each other. **d)** UMAP clustering: **i)** of tumors relates to organ systems rather than patient survial, based on mRNA expression data for 20530 genes in 9112 patients across 32 cohorts of cancer types (**Table 3.1**). However, **ii)** Scaling-informed machine learning (SIML) reduces the genes to just those that scale with *LMNB1* in 17 cancer types, and these are then color coded as 'regulated' which also includes 2 more cancers for which transcription factor *FOXM1* scales with *LMNB1* expression. **iii)** The 'regulated' tumors show significantly longer median patient survival.

3.3.2 Expression scaling across tumors and adjacent-uninvolved tissues

The observed co-clustering of system-related organs (e.g. digestive system) (**Fig.3.1d-i,S3.1a**) implies a strong influence of normal tissue gene expression, and so we start with scaling analyses of all 16 cancer cohorts in TCGA (of the 32) that provide expression data for both adjacent-uninvolved patient tissue and tumor (**Table 3.1**). Given that filaments of lamin-B1 assemble quickly around chromatin after mitosis (**Fig.3.1a**), the physics-motivated power law

$$[\text{mRNA expression of gene}] \sim [\text{LMNB1}]^{\beta_{\text{gene}}} \qquad \text{Eq.1}$$

was first applied to a marker of proliferation, *MKI67* (**Fig.3.2a**). Representative tumors show lung and breast (not liver) exhibit a reasonably continuous trend from adjacent normal to tumor over a ~2^6 range of apparent expression. All three tumor types show reasonable scaling with exponents from ~0.8 to ~1.3 (R^2 = 0.6-0.8). The lack of scaling for normal liver and the very low *MKI67* suggests little to no proliferation.

Similar analyses for *COL1A2* versus *COL1A1* (**Fig.3.2b**) yield exponents of a_{COL1A2} = 0.83 ± 0.11 (S.D.) when fit with:

$$[\text{mRNA expression of gene}] \sim [\text{COL1A1}]^{\alpha_{\text{gene}}} \qquad \text{Eq.2}$$

Although the scaling is slightly lower than expected of stoichiometric scaling (a = 1), the R^2 > 0.84 suggests co-regulation over a large range of apparent expression (i.e. ~2^{10} fold, which is much broader than *LMNB1*). *COL4A2* versus *COL4A1* was also examined because this main basement membrane collagen shares a promoter (Fig.1b) (unlike collagen-1), and scaling exponents are only ~12% lower than stoichiometric scaling (R^2 = 0.84 to 0.95) (Fig.S2a). Basement membrane *COL4A1* does not scale well with fibrous *COL1A1* (Fig.S2b), consistent with these ECM's being distinct.

3.3.3 Heatmaps of high Cell Cycle, LMNB, and Collagen-I differ from varied Cytoskeleton-LMNA

Conventional heatmaps of (tumor/adjacent) (**Fig.3.2c-f**) always show higher *COL1A1* in tumors as well as higher B-type lamins (*LMNB1* and *LMNB2*) and higher cell cycle plus DNA-repair genes for progression through cell cycle checkpoints(Dasika *et al.*, 1999). Protein changes for liver (tumor/adjacent) conducted by quantitative Mass Spectrometry (MS) proteomics analyses (n=3 patients) show very good concordance with TCGA transcriptome trends. Structural proteins are generally abundant and well-suited for accurate quantitation. Overall, >85% of proteins and RNA showed concurrent upregulation/downregulation (**Fig.3.2f**), and *LMNB1* and *LMNA* were up in both analyses as were all three of the MS-detected, cell cycle related DNA repair factors.

Unlike *LMNB1*, *LMNA* is lower than adjacent in about half the tumors, and shows roughly similar variability to some of the cytoskeleton and adhesion structure genes. To relate such observations to a limited literature, we focus on liver and its differences from lung and breast cancers (R^2 = 0.4; **Fig.S3.2c**). Liver shows relatively higher basement membrane ECM (e.g. collagen-4), adhesions-cytoskeleton, and also *LMNA,* -- unlike lung and breast cancer. Low Lamin-A/C in lung(Kaspi *et al.*, 2017) and breast(Capo-chichi *et al.*, 2011) cancers potentially facilitates invasion and growth based on knockdown studies *in vitro* and *in vivo* with a lung cancer line (Harada *et al.*, 2014b). In various 2D cultures, LMNA promotes adhesive spreading and actomyosin assembly (Swift *et al.*, 2013) downstream of the Serum Response Factor (SRF) pathway(Ho *et al.*, 2013; Buxboim *et al.*, 2014), and high *LMNA* in a few cancers including liver cancer (**Fig.3.2e**) favor co-clustering with *SRF* co-activators *MKL1, MKL2* and/or their target nonmuscle myosin-IIA (*MYH9*)(Buxboim *et al.*, 2014). Paxillin (*PXN*) is very similar as are basement membrane collagen-4's (unlike collagen-1), but so is *YAP1* in the Hippo pathway for growth(Dupont *et al.*, 2011), differing from *SRF* and its target adhesion factor vinculin (*VCL*)(Miralles *et al.*, 2003). Overall, the results are consistent with the known matrix mechanosensing pathway of LMNA-contractility-adhesion downstream of basement membrane density (or stiffness) (**Fig.3.2e-sketch**), but importantly this pathway segregates from *LMNB's.*

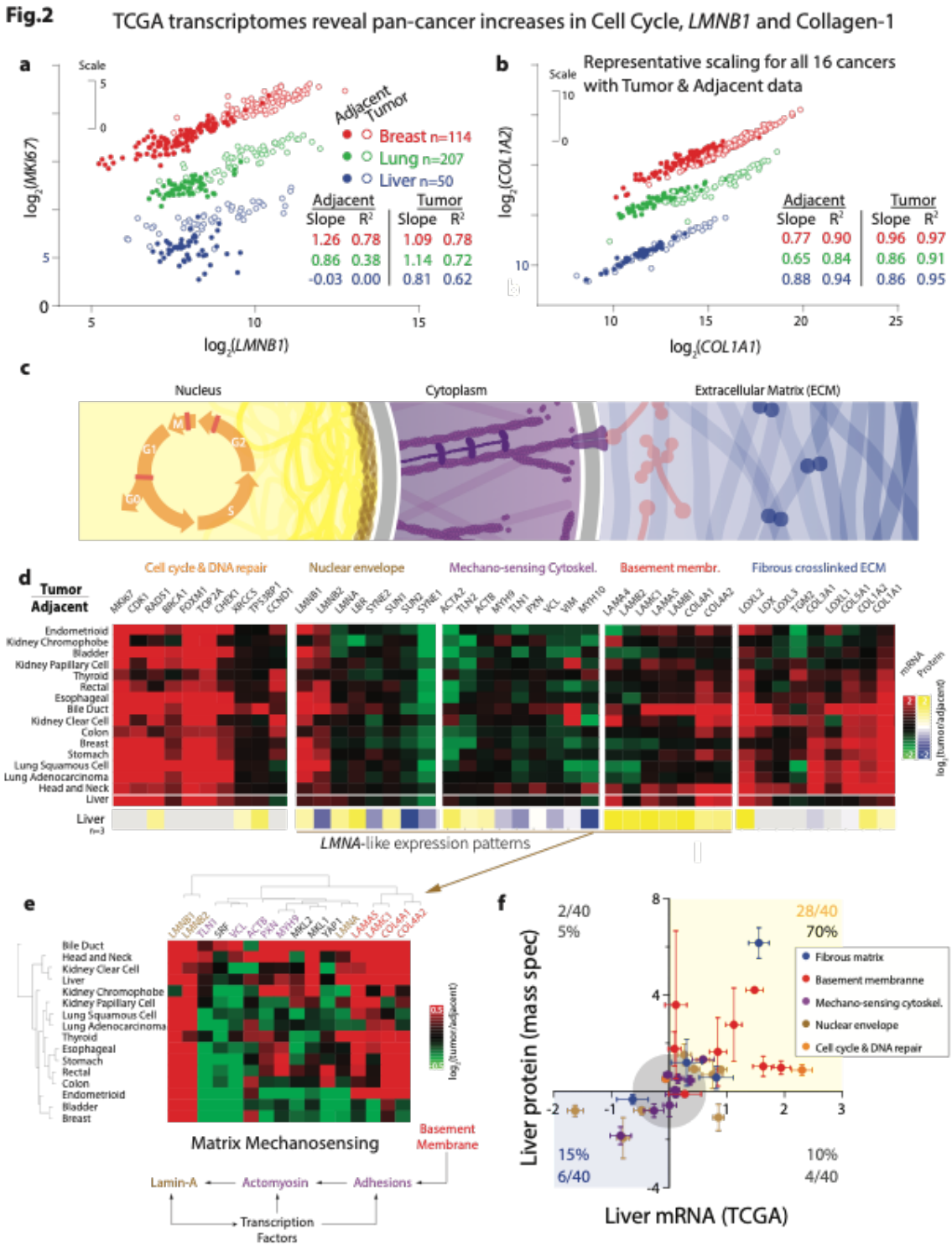


Figure 3-2 TCGA transcriptomes reveal pan-cancer increases in Cell Cycle, *LMNB1*, and Collagen-1

a) RNA reads (RSEM) for proliferation marker *MKI67* versus nuclear lamina factor, *LMNB1*, reveals scaling in all three tumor types (R^2 > 0.62) but not adjacent Lung Adenocarcinoma or Liver tissue. For clarity, Lung Adenocarcinoma and Breast data are shifted up 5 and 10 units, respectively, **b)** RNA reads (RSEM) for the two subunits of the collagen-I heterotrimer scale together across all three cancers and corresponding adjacent tissue (R^2 > 0.87). For each patient, tumor and adjacent tissue RNA are reported, and tumor shows more collagen-I on average. For clarity, Lung Adenocarcinoma and Breast data are shifted up 3 and 6 units, respectively **c)** Schematic polymer systems of main interest. **d)** Standard heatmap of $\log_2$(fold-change of RNA in tumor relative to adjacent tissue) averaged over all patients for the 16 tumors having >4 patients with adjacent tissue data. Yellow-blue heatmap of $\log_2$(fold-change of protein in tumor relative to adjacent tissue) as obtained in LFQ units from mass spectrometry (MS) for n=3 Liver Cancer patients. **e)** RNA changes for various matrix mechanosensitive factors including Lamin-A across the 16 cancers; such factors do not vary with either collagen-I or lamin-B. **f)** Ratios of (tumor/adjacent) for Liver protein and RNA show 85% of data in the first and third quadrants.

3.3.4 Proliferative genes scale with *LMNB1* and predict poor survival

In analyzing all 32 cancer cohorts, scaling with *LMNB1* across patients with a given tumor type is evident for some cancers (such as liver) for well-known cell cycle genes *TOP2A*, *FOXM1* and *MKI67* (**Fig.3.3a-b,S3.2a**, n=371). The results are consistent with the smaller dataset (**Fig.3.2a**, n=50). *TOP2A* is among the top-24 most upregulated genes (tumor versus adjacent normal) together with fifteen other genes (many associate with proliferation) that also scale well with *LMNB1* (**Table 3.2**). The anti-correlation for the hepatocyte-expressed, complement-related factor *MASP2* (**Fig.3.3c**) is consistent with a shift away from differentiation given that such genes accumlate in a protracted cell cycle(Palozola *et al.*, 2017). Anti-correlations differ in lung cancer (**Fig.S3.2b**), consistent with a distinct lineage.

Scaling exponents β_{gene} and R^2 of the fits for all ~20k expressed genes (per Eqs.1,2) yield a 'sideways-volcano plot' (**Fig.3.3d**), and the most well-correlated genes (i.e. β_{gene}>0.5, R^2>0.5) are proliferation-related (83% of 168 genes for liver cancer). To assess whether *LMNB1* transcript levels relate to the number of copies of the *LMNB1* gene, genomic data for each liver cancer patient was analyzed(Mermel *et al.*, 2011), noting that copy number variations are consistent with malignancy and can predict patient survival(Pfau and Amon, 2012; Sankaranarayanan *et al.*, 2015; Aiello and Alter, 2016; Bradley *et al.*, 2019). Transcript levels binned on gene copy number fit a power law of ~1 (**Fig.3.3e**) – although patients with one copy of the *LMNB1* gene are bimodal distributed and only the low-expressors fit the trend. Consistent with gene dosage DNA → mRNA

→ protein, our proteomics for liver cancer detected LMNB1 plus three proliferation factors that are all upregulated in tumors (**Fig.3.2d,e**). Experiments in cancer cells below will directly test whether *LMNB1* DNA levels indeed affect proliferation.

Scaling with respect to *MKI67* yields a similar number of genes that are >95% identical to the *LMNB1* gene set (**Fig.S3.2c**), underscoring invertibility of the scaling, whereas fits on linear scales (i.e. Pearson) yield <30% of these genes. When repeated for all 32 cancers, 25 show the mitotic factor *KIF20A* is the gene that scales most often with *LMNB1*, and a total of 866 unique genes also scale across 27 of the 32 (with 5 showing no scaling genes), with only 242 genes appearing in more than 1 tumor (**Fig.3.3f**). The number of genes that are shared across multiple cancers clearly decreases with the number of cancers considered, but the theoretical solution to the maximum number of common set of genes is an open problem in mathematics known as the Maximum k-Subset Intersection (MSI) problem which lacks even approximate solutions(Xavier, 2012). To compute, for example, the maximum number of strongly scaling genes that overlap across 17 of the 32 cancers, the brute force method would be to consider all $^{32}C_{17}$ = 565,722,720 combinations.

For an efficient computation in a heurtistic approach, we represented the genes in a matrix populated with indicator variables (per **Supplementary Theory: Gene Overlap**). For *LMNB1* scaling genes, it so happens that tumors that show maximum overlap of genes for a particular number of tumors (n) being considered, are also present in the group that shows maximum gene overlap in n+1 tumors with the addition of another tumor. We listed the cancer which, when added to a growing group of cancers, gives the maximum number of genes that overlap (**Fig.3.3g**). Based on this approach, for example, the maximum overlap of genes obtained from all combinations of choosing 17 tumors is 25 genes. Notable among the 25 genes that scale with *LMNB1* is *FOXM1* (**Fig.3.3e inset**), which is the only transcription factor and a regulator of cell cycle(Chen *et al.*, 2013; Grant *et al.*, 2013). *FOXM1* scales strongly with *LMNB1* in the 17 cancers plus 2 more (thymoma and a lymphoma), and all were denoted as the “regulated” cancers among TCGA’s 32 cancers (per

Fig.3.1d-ii). *TOP2A* is also a notable mitotic gene regulated directly by FOXM1 (Nielsen *et al.*, 2020).

The biological significance of these scaling results is best assessed by patient survival. For liver cancer (371 patients), median survival is ~3-4 yrs for high *LMNB1* patients but almost twice as long for low *LMNB1* (**Fig.3.3h**). This partitioning is independent of etiology including alcohol or hepatitis (**Fig.S3.3d**). Across all genes, the median survival for high/low expressors shows a significant hazard ratio for 3,464 genes, and among the genes that predict poor survival when expression is high (2,111), >95% of the genes scale with *LMNB1* (**Fig.3.3i**). Staging of primary tumors in terms of size and invasion also shows *LMNB1* increasing (**Fig.S3.2e**), whereas *LMNA* and *COL1A1* show no trends. A sideways-volcano plot for *LMNA* shows only one gene that scales strongly (**Fig.S3.2f**), which underscores the distinctive significance of *LMNB1*. Across the 32 cancers, *LMNA* shows strong scaling genes in very few cancers, with only 75 genes appearing in >1 cancer, dropping down to a maximum overlap of 1 gene in groups of 4 tumors (**Fig.S3.3g**), and does not predict patient survival (**Fig.S3.3h**). More generally, among the 17 cancers that are regulated with 25 overlapping genes or more scaling with *LMNB1* (**Fig.3.3g**), patients with high expression of *LMNB1, FOXM1, TOP2A* are predicted to have poor survival in 7-9 cancers (**Fig.3.3j**). Only thymoma shows pro-survival with high *LMNB1* and high *TOP2A*, which might reflect a role for Lamin-B1 in thymus development of T-cells(Yue, 2017; Wang *et al.*, 2020). In contrast, only 1-2 cancers among the remaining 15 cancers show high expression of *LMNB1, FOXM1, TOP2A* has any significance in predicting (poor) survival, which highlights the regulated versus dysregulated distinction even for well-accepted cell cycle markers - *FOXM1, TOP2A*.

Lamin-B2 is farnesylated like lamin-B1 but unlike mature lamin-A, and tight association of both LMNB's with lipid in the nuclear envelope implies scaling with chromatin mass (**Fig.3.1a**). Indeed, recent studies of proliferating, embryonic cardiomyocytes suggest *LMNB2* facilitates attachment of microtubules to centromeres(Han *et al.*, 2020). In our scaling analyses, *LMNB2* like *LMNA* quickly drops down to a maximum overlap of 1 gene in groups of 8 tumors, but genes that scale strongly with *LMNB2* can predict poor survival similar to *LMNB1* (**Fig.S3.3i, S3.4a,b**).

Moreover, the cell cycle transcription factor *FOXM1* also scales with *LMNB2* as with *LMNB1* (**Fig.S3.4a,b**), and thus further motivates experiments at for FOXM1 regulation of *LMNB1*.

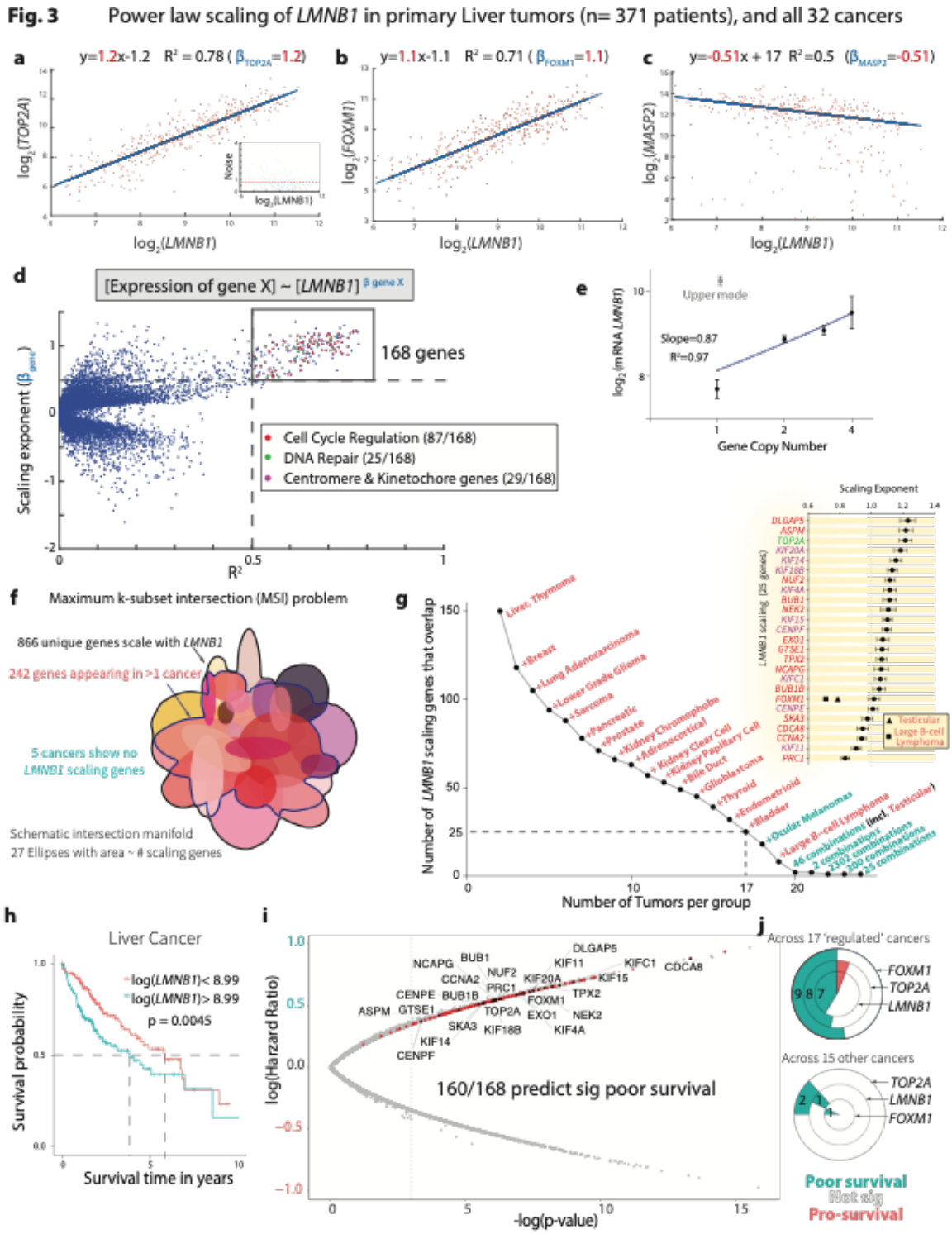


Figure 3-3 Power law scaling of *LMNB1* in Liver tumors (n=371 patients), and all 32 tumors

a-c) Plots versus *LMNB1* RNA for correlated cell cycle genes *TOP2A* and *FOXM1* as well as for anti-correlated gene *MASP2*. The R^2 and scaling exponent (β_{gene}) are indicated. **d)** Genes that scale with *LMNB1* RNA. **i)** For all 17,958 genes, β_{gene} versus R^2 gives a sideways-volcano plot; 168 genes scale strongly β_{gene}> 0.5 and well R^2 > 0.5, and most relate to mitosis. **e)** Gene copy numbers of *LMNB1* in liver cancer patients is almost linear in RNA expression, although patients with a single copy are bimodally distributed and only the lower mode fits the trend. **f)** Schematic Venn diagram of overlapping gene sets that scale with *LMNB1* in 32 cancers. **g)** *LMNB1* strong scaling genes show overlap across tumors, decreasing as the number of tumors being considered increases (see **Gene Overlap**). *Inset*: A maximum of 25 genes scale with *LMNB1* across 17 tumors. The 19/32 tumors where transcription factor *FOXM1* scales with *LMNB1*, denoted as 'regulated' tumors include the 17 tumors plus Testicular and Large B-cell Lymphoma. **h)** For patients with *LMNB1* RNA levels that exceed the respective median levels for all patients, the median time for survival is significantly shorter ($p < 0.05$) by 2-3 yrs in Kaplan Meier (KM) plots. **i)** For all genes, similar KM analyses are summarized by the hazard ratio of the two cohorts plotted against the p-value, yielding 3,464 genes that show significant differences, including 161 (of 168) genes that scale with *LMNB1* and show poor survival when expression exceeds the median. **j)** High expression of *FOXM1, TOP2A, LMNB1* predicts poor survival in 9,8, or 7 of the 17 tumor types, with the 7 cancers being: Liver cancer, Pancreatic Cancer, Adrenocortical Carcinoma, Lower Grade Glioma, Sarcoma, Kidney Clear Cell and Kidney Papillary Cell. Among the other 15 cancers, only Melanoma for *LMNB1*, Mesothelioma for *FOXM1*, and both cancers for *TOP2A* show significantly poor survival in patients with higher expression

3.3.5 *FOXM1* directly regulates *LMNB1* expression, but perturbing LMNB1 perturbs cancer cell cycle

Chromatin-immunoprecipitation with anti-FOXM1 followed by sequencing (ChIP-Seq) shows FOXM1 binds the promotor regions of *LMNB1* as well as cell cycle genes *TOP2A, KIF20A, KIF11* in two cell lines (see Methods). No signals were detected for either *LMNA* or the constitutive heat shock gene *HSP90AA* (**Fig.3.4a**). To test *LMNB1* gene regulation by FOXM1, we designed a promoter-reporter plasmid and transfected it into A549 lung adenocarcinoma cells and U2OS osteosarcoma cells for detection and perturbation of GFP reporter signal (**Fig.3.4b,c**). Siomycin reduces FOXM1 expression and promotes its degradation (Radhakrishnan *et al.*, 2006; Bhat, Halasi and Gartel, 2009), and FOXM1-i inhibits DNA binding at the consensus sequence(Gormally *et al.*, 2014). We find that cells treated with Siomycin (for more than 2 hr) and FOXM1-i (at ~6 hr) show less GFP expression, consistent with reduced *LMNB1* promotor binding (**Fig.3.4d,e**). FOXM1 inhibition decreases mitotic cell numbers (by ~6 hr) as expected(Laoukili *et al.*, 2005). FOXM1 inhibition suppresses Lamin-B1 protein and TOP2A (**Fig. 3.4f,g**), whereas mitotic cell accumulation induced with Nocodazole (18 hr) increases Lamin-B1 (**Fig. 3.4f**), as also occurs with FOXM1(Laoukili *et al.*, 2005). The results are consistent with simple cell cycle trends (**Fig. 3.1c**).

Anti-FOXM1 immunofluorescence and *LMNB1* gene-edited with red fluorescent protein (RFP) both show nuclear localization in interphase and chromatin association in mitotsis (**Fig. 3.4h**), and FOXM1 inhibition decreases anti-FOXM1 intensity as expected. Intensity analyses further show FOXM1 increases through cell cycle (**Fig. 3.4j**), with many G0/G1 cells showing no FOXM1, consistent with its degradation at mitotic exit(Laoukili *et al.*, 2008; Park *et al.*, 2008). Importantly, Lamin-B1 increasing linearly with FOXM1 (**Fig. 3.4k**) is in excellent quantitative agreement with the scaling analyses of 17 regulated cancers in TCGA (**Fig. 3.3g**).

Lamin-B1 dynamics differ from FOXM1 in that it disassembles and is inherited by daughter cells(Gerace and Blobel, 1980), remaining non-degraded and non-zero throughout cell cycle as is confirmed by imaging live cells and fixed cells (**Fig. 3.4k,5a,b**). Note that the gene-edited RFP-*LMNB1* and likewise GFP-histone-H2B use the endogenous promoters to drive expression in the lung cancer cell line A549(Pfeifer *et al.*, 2018), which avoids potential artifacts of overexpression. Power law increases in RFP-LMNB1 and GFP-histone-H2B intensities per nucleus up to mitosis, allow one to to eliminate the time dependence in gene-gene scaling with cell cycle at the single cell level (per **Fig. 3.1c**; and see **Theory Supplement**). This is important given tumors have differences in mean doubling time and is consistent with scaling between cancer patients (**Fig. 3.3a-d**).

RFP-LMNB1 and anti-LMNB1 intensities not only scale with total DNA-staining intensity per nucleus, including high ploidy cells relevant to cancer (**Fig. 3.5c,d**), but *LMNA* knockdown has no effect. This is consistent with the absence of *LMNA* from the *LMNB1*-scaling set with (**Fig. 3.3d**), and with distinct roles for lamin genes in mechanosensing (*LMNA*) (**Fig. 3.2f**) versus proliferation (*LMNB*'s). To assess effects of LMNB1 on cancer cell proliferation, we transfected GFP-LMNB1 plasmid or shLMNB1 plasmid into U2OS cells, and pulsed incorporation of the nucleotide analog EdU (for 1hr) together with DNA staining to quantify cell cycle stage(Pfeifer *et al.*, 2018). High EdU signal indicates ongoing DNA synthesis (S-phase), and low EdU signal indicates G1 or G2 phases, depending respectively on low or high DNA intensity, i.e. '2N' or '4N' (**Fig. 3.5e**). Knockdown cells show more cells in G1 compared to wildtype or overexpressing cells (identified by GFP signal at the single-cell level), whereas the latter were more in S-phase and G2 phase (**Fig. 3.5f**), implying

low *LMNB1* suppresses cancer cell cycle (**Fig. 3.5g**). These cancer cell line results agree with fibroblast results(Shimi *et al.*, 2011), and imply *LMNB1* is a novel knockdown target to suppress cancer growth.

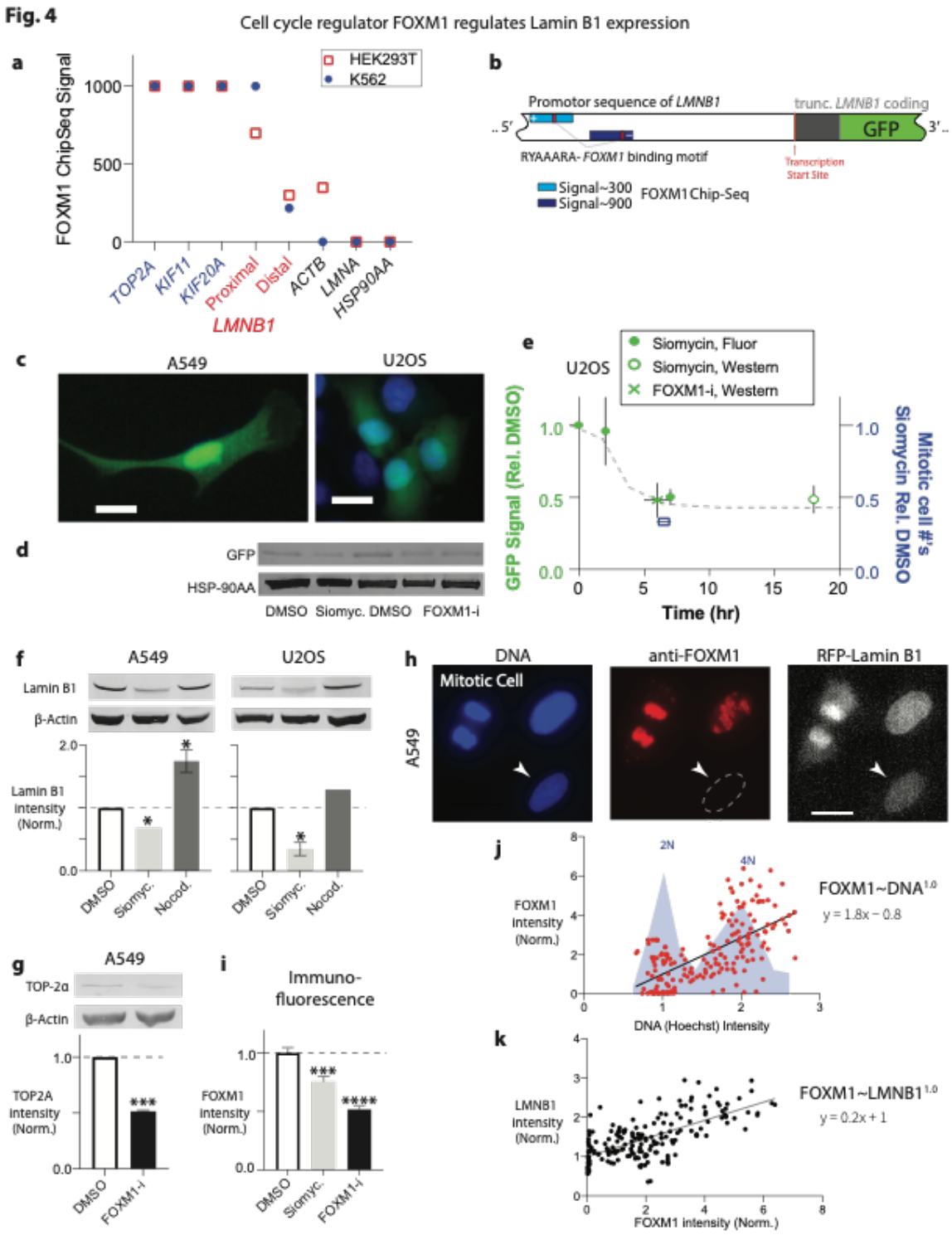


Figure 3-4 FOXM1 transcription factor regulates Lamin-B1 expression

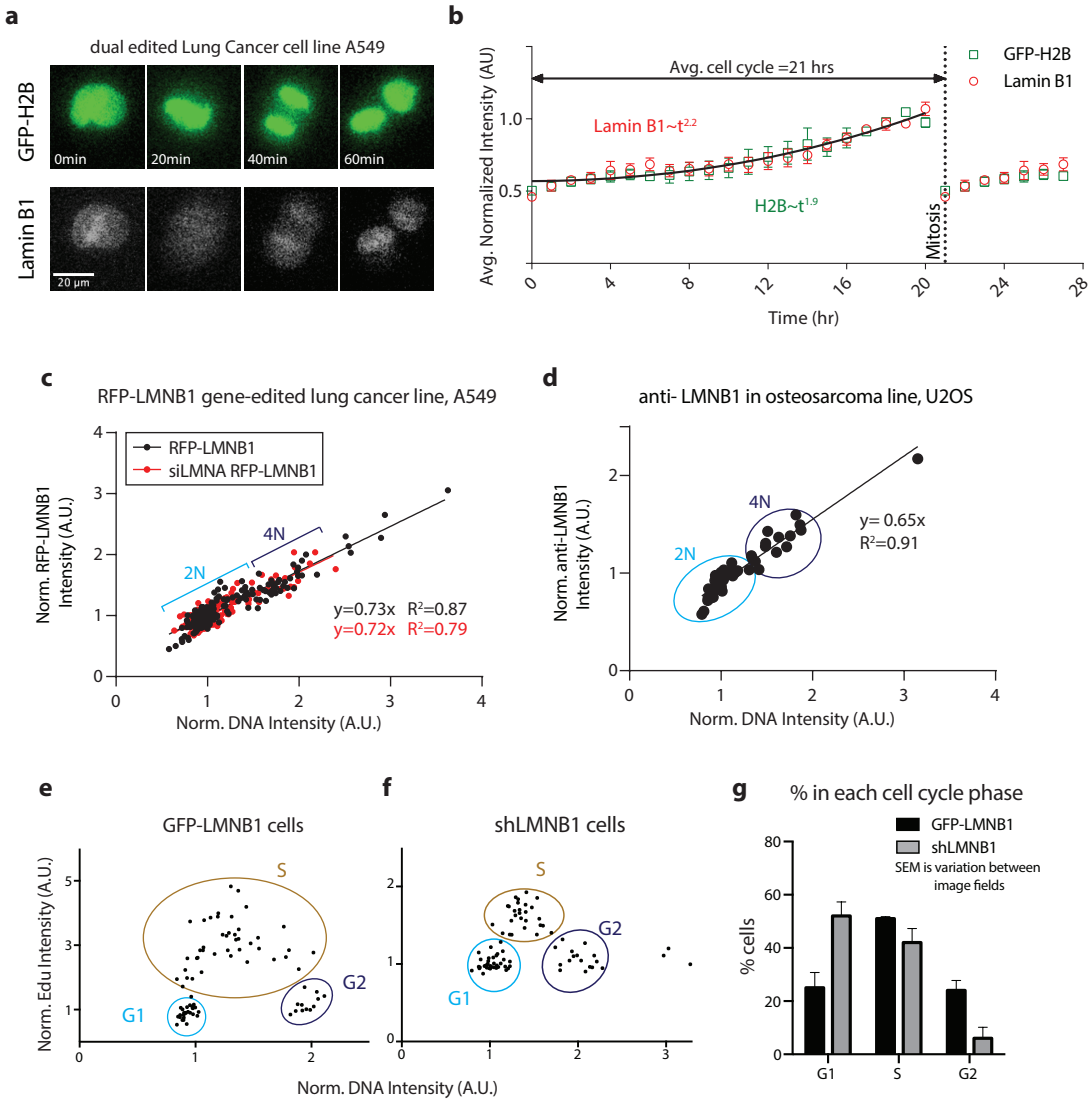
a) Chip-Seq analyses show FOXM1 binds to the promotor regions of cell cycle regulated genes including *TOP2A, KIF11, KIF20A*, and two sites in *LMNB1*. No binding is evident in *LMNA* and *HSP90AA*. **b-e)** GFP promoter-reporter for assessing regulation of *LMNB1*, and its transfection into A549 and U2OS cancer cell lines shows decreased GFP in immunoblots of cells treated with FOXM1 inhibitors Siomycin for 18 hr and FOXM1-i for 5-7 hr (representative of n=3). Additional analysis of GFP in images of U2OS cells normalized to DNA, and again shows reduced GFP expression after 7 hr of Siomycin, which also reduces mitotic cell counts as expected (Avg. +/- SEM for >50 cells measured). **f,g)** Lamin-B1 protein is also decreased by FOXM1 inhibition (Siomycin, 18 hrs) as is TOP2A (5hrs), but enriching for mitotic cells with Nocodazole increases Lamin-B1. **h-k)** Mitotic and G2-phase A549 cells show high intensities of stained DNA, anti-FOXM1, and gene-edited Lamin-B1, in contrast to G1-phase cells with low to negligible anti-FOXM1 intensity. FOXM1 inhibition (6 hr) reduces anti-FOXM1 intensity as expected, but FOXM1 increases in cell cycle as DNA intensity doubles. FOXM1 and LMNB1 scale linearly at protein levels through the cell cycle.

Figure 3-5 LMNB1 scales with DNA in vitro and LMNB1 levels modulate cell cycle as a proto-oncogene

a,b) Live cell imaging of gene-edited lung cancer line expressing GFP-H2B and RFP-LMNB1 shows parallel increases in intensities normalized to telophase in mitosis, with a mean cell cycle of 21±0.35 hrs. **c)** Intensity of RFP-LMNB1 is linear versus Hoechst-stained DNA in fixed A549s, even with lamin-A knockdown. Intensities are normalized to cells in the non-replicated state, '2N'. **d)** anti-LMNB also shows LMNB increases linearly with DNA in osteosarcoma-derived U2OS cells. **e-g)** EdU incorporation (1 hr) in replicating cells in combination with Hoechst-stained DNA identifies cell cycle stage (G1/S/G2). Knock down with shLMNB1 in U2Os cells reveals a smaller percentage of cells proceeding to S and G2 compared to overexpressing GFP-LMNB1 cells. Error bars indicate SEM values across image fields

Fig.5

LMNB1 scales with DNA in vitro and LMNB1 levels modulate cell cycle as a proto-oncogene



3.3.6 Pan-cancer exponents for genes scaling with *LMNB1, ACTA2, COL4A1, COL1A1*

The 'universality' of power law scaling exponents across cancers is illustrated by results for distinct sets of 17 cancers (from the 32) for *LMNB1* (**Fig. 3.3e**, **Table 3.1**) and a few other genes, noting also that 17 is a slight majority of the 32 total tumor types in TCGA. *LMNB1* scaling genes yielded 25 genes, that are all cell cycle related including *FOXM1*, with exponents across tumors that range from 0.83-1.23 (**Fig. 3.6a**). Higher values (>1) are consistent with a lamina (~Area) that surounds replicating DNA (~Volume) (**Fig. 3.1a**) or perhaps a bias from early or late cell cycle genes (**Fig. 3.1c**). In comparison, all four genes that scale with the cytoskeleton gene *ACTA2* across 17 cancers

have super-stoichiometric exponents (>1), and three are also cytoskeleton while one is a membrane calcium regulator (**Fig. 3.6b**).

Four genes also scale with *COL4A1* for a different set of 17 cancers, and again show a range of power laws for these basement membrane ECM genes (**Fig. 3.6c**), including a pan-cancer exponent for *COL4A2* of 0.92 ± 0.02 within 10% of expected results for a shared promoter (**Fig. 3.1b**). Ten fibrillar ECM or membrane genes scale with *COL1A1* (**Fig. 3.6d**), including *COL1A2* as noted previously (**Fig. 3.2b**). Membrane factors include one integrin and also *FAP* (fibroblast activating protein) which is a membrane-bound gelatinase targeted in the clinic(Hofheinz *et al.*, 2003). As with *COL4A1,* all *COL1A1* exponents are sub-stoichiometric (i.e. <1), consistent with these major ECM components serving as scaffolds for the other factors to decorate like leaves on a tree (**Fig. 3.1a**). Equally important, scaling of *COL5A1* ~ $COL1A1^{0.84\text{-}0.92}$ from TCGA for 17 cancers (and across all 32 cancers in TCGA $\alpha = 0.84 \pm 0.02$, SEM) agrees with mass spectrometry-derived scaling proteomic results(Swift *et al.*, 2013) based on COL1A1 ~ $E^{1.5}$ and COL5A1 ~ $E^{1.3}$ giving COL5A1 ~ $COL1A1^{0.87}$. Exponents for other genes agree on average within 20%, and agreement of tumor and normal transcript and protein is evidence of 'universality' in scaling.

3.3.7 Single-cell RNA-seq with Scaling-Informed Machine Learning (SIML)

Bulk sequencing of tumors mashes together a diversity of cell types at different stages of cell cycle, and thus motivates a scaling analyses of single-cell RNA-seq data (scRNA-seq). Liver cancer biopsies from chemotherapy-treated patients have only recently been characterized by scRNA-seq and show some malignant hepatocellular carcinoma cells (HCC) among many types of stromal and immune cells (Ma *et al.*, 2019). UMAP clusters of the specific cell types show cancer associated fibroblasts (CAFs) have high reads of *COL1A1, COL1A2, FAP, ACTA2, CNN1* (**Fig.3.6e,S3.5**). However, *COL1A1* and all of its pan-cancer scaling genes (**Fig. 3.6d,S3.5**) are high in a sub-population of CAFs distinct from those expressing *ACTA2* and its bulk-scaling genes (**Fig. 3.6b, S3.5**)**.** Fibroblasts with high expression of FAP have been reported to be distinct from those with high expression of ACTA2 (smooth muscle actin protein, αSMA) in cancer associated stroma from cancer patients (Tchou *et al.*, 2013; Kilvaer *et al.*, 2015; Öhlund *et al.*, 2017), tumor xenografts

(Özdemir *et al.*, 2014) and *in vitro* experiments (Avery *et al.*, 2018), but across cancers what genes best distinguish the two CAF phenotypes has been uncertain. Power law scaling thus discriminates synthetic CAFs (*COL1A1, COL1A2, FAP...*) from contractile CAFs (*ACTA2, CNN1...*), even though all are fibroblast-clustered 'fibrotic' genes. SIML provides a new and general method to better define and resolve closely related lineages.

LMNB1 scRNA-seq reads are dominated by the easily extracted (non-adherent) Tcells (**Fig. 3.6f-*i*; Fig.S3.6a,b**) even though cancer cells such as A549s clearly express *LMNB1* (**Fig. 3.6f-*ii*; Fig.S3.6c**). In diverse tissue cell populations as well as in a related population of cells, high reads of *LMNB1* are coincident in cell populations with high reads of *FOXM1* (**Fig.S3.6b,c**) indicative of *LMNB1* expression regulation (**Fig. 3.4**). Scaling exponents for *LMNB1*-scaling genes that were averaged for liver cancer and A549 scRNA-seq nonetheless correspond well to those from TCGA's pan-cancer exponents (**Fig. 3.6d**). Deviations likely result from low reads from *LMNB1* and other mitotic genes. Collagens are ~10-fold more abundant, and CAFs and endothelial cells express *COL4A1, A2* isoforms that scale with exponent = 0.83 versus 0.86 for A549s (**Fig. 3.6f-*iii,iv***) and 0.92 for bulk from TCGA (**Fig. 3.6c**). *ACTA2* scaling gene *CNN1* shows weaker scaling in scRNA-seq (**Fig. 3.6f-*iv***) by ~20% than the pan-cancer exponents in bulk TCGA, which is about typical for deviations. Nonetheless, the raw reads based scRNA-seq scaling analyses seem reasonably consistent with 'universality' in scaling.

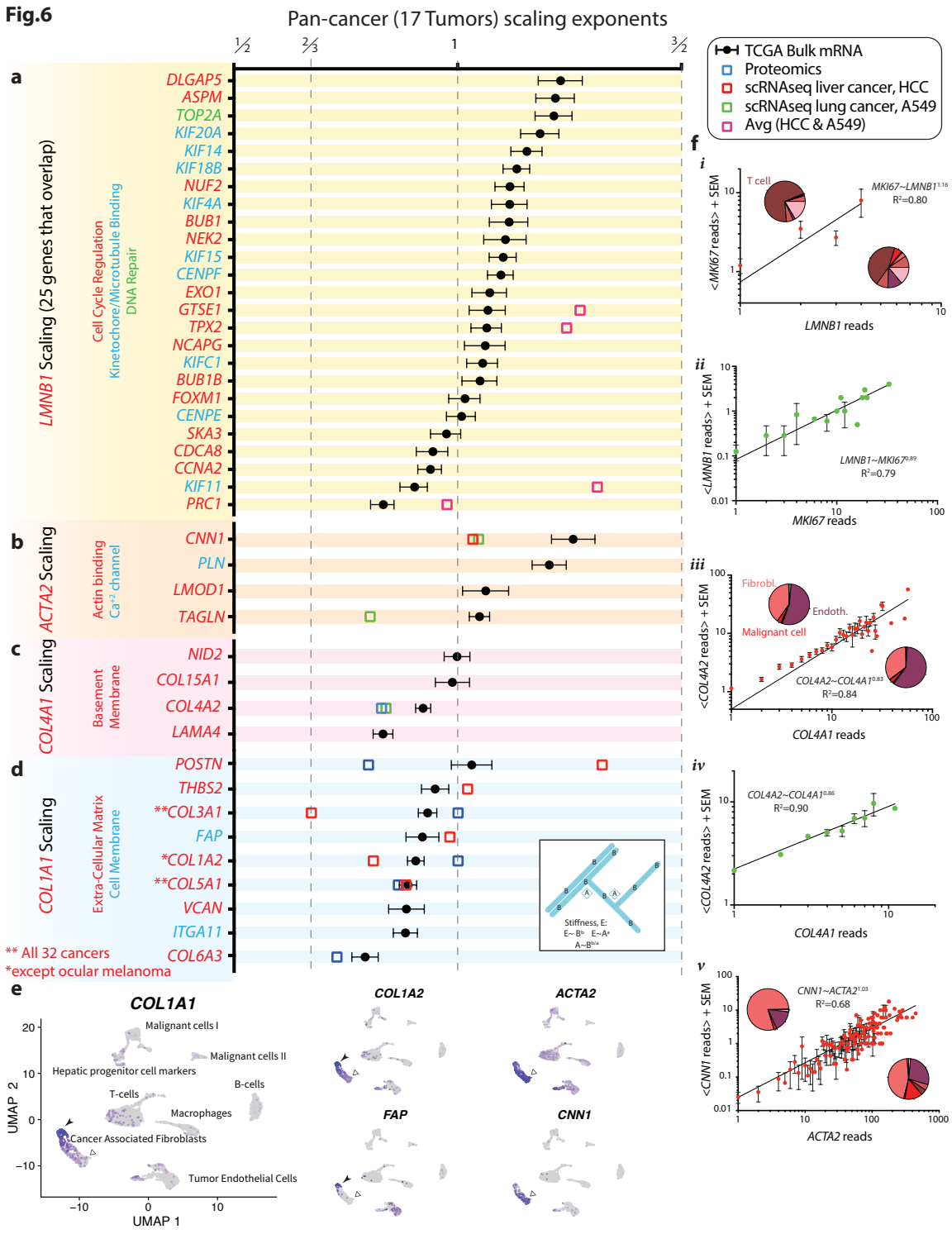


Figure 3-6 Pan-cancer power law exponents

a-d) (circle) Exponents for genes in TCGA that show indicated scaling with slope > 0.5 and R^2 > 0.5 across 17 primary tumor cancer types (See **Gene Overlap** Supplement). (blue square) Exponents obtained from proteomics across normal tissues and model tumors (blue square).

Exponents from single-cell RNA-seq: (red square) hepatocellular carcinoma, HCC; (green square) lung cancer derived A549 line; (pink square) average exponent of HCC and A549. For *LMNB1*, 25 genes show maximum overlap in only 1 grouping of 17 tumors. For *ACTA2*, 2 combinations of 6 genes show maximum overlap in groups of 17 tumors; and 4 genes that appear in both combinations. For *COL4A1*, 21 combinations of 7 genes show maximum overlap in groups of 17 tumors; 4 genes that appear in >18 combinations. For *COL1A1*, 11 combinations of 9 genes show maximum overlap in groups of 17 tumors; 9 genes that appear in >6 combinations.**e)** Machine-learned, dimensionally reduced projection (UMAP) of single-cell mRNA sequencing of liver cancer (HCC) labeled with standard marker identified lineage. Each point is a cell, with purple indicating detected expression levels of *COL1A1, COL1A2, ACTA2, FAP, CNN1* **f)** Power law fits of raw reads from single-cell mRNA sequencing: **i-ii)** *LMNB1* vs *MKI67* for HCC or A549's. For HCC, ~45% of cells with non-zero reads of *LMNB1* or *MKI67* are Tcells, and <5% are malignant cells. **iii-iv)** *COL4A1* vs *COL4A2* from HCC patients or A549 cells. Pie charts for HCC: <5% of malignant cells have non-zero reads for *COL4A1* or *COL4A2*, versus ~60% of Endothelial Cells, ~35% are Fibroblasts. **iii)** *ACTA2* vs *CNN1* from HCC scale together as myofibroblastic genes and show similar cell-type specific expression profiles.

3.3.8 ECM fibrosis genes scale with *COL1A1* across cancers and predict liver cancer survival

Lastly, because fibrotic collagen-1 and cell cycle lamin-B1 are both high in all 16 cancers for which tumor and adjacent TCGA data is available (**Fig. 3.2d**), scaling with *COL1A1* was scrutinized, starting with primary liver cancer (n=371). *COL1A1* expression varies ~16,000-fold (versus ~32-fold for LMNB1 in **Fig. 3.7a**), and does not relate to genome copy number changes in *COL1A1* (**Fig.S3.7a**), consistent with expression by non-malignant stromal cells (found also in adjacent and normal tissue). First, even though fibrous *COL1A1* and most associated ECM are abnormally high within a given patient across all cancers (**Fig. 3.2d**), between patient tumors these genes do not correlate with *LMNB1* (**Fig. 3.7a**: α_{LMNB1} ~0 and R^2 < 0.1). Once again, *COL1A2* scales almost linearly with *COL1A1*: α_{COL1A2} = 0.86 (**Fig. 3.7b**), and an ECM protease gives α_{MMP2} = 0.79 (**Fig. 3.7c**). *COL1A2* scaling also has the highest R^2 among all ~20k genes. Among 162 genes that scale with *COL1A1* (α_{gene} > 0.5, R^2 > 0.5), the majority are ECM genes (**Fig. 3.7d**). *COL4A1* and *COL4A2* fit well but with weak scaling (α_{gene} < 0.5, R^2 > 0.5) suggestive of a (COL1 / COL4) ~ (bulk/surface) scaling (**Fig. 3.1a**). The only other transcripts with R^2 > 0.9 are *COL5A1* and *COL3A1* (**Fig. 3.7d**), and these associate with collagen-1 fibers(Gelse, Pöschl and Aigner, 2003), scale with *COL1A1* in proteomics(Swift *et al.*, 2013), and scale 'universally' across all 32 cancers (**Fig. 3.7d; Gene Overlap**).

To assess survival of cancer patients in relation to 'fibrotic' genes, we focused first on the 16 (tumor/adjacent) cancers. This is because, unlike *LMNB1* which is expressed by malignant cells

(and predicted poor survival in 7 cancers including liver cancer, **Fig. 3.3g,i**), *COL1A1* and *COL1A2* are expressed primarily by fibroblasts found in both tumor (i.e. CAFs) and adjacent tissues (**Fig. 3.6e**). Most scaling genes associated with CAFs (i.e. *ACTA2* and *COL1A1* gene sets per **Fig. 3.6b,d**) predict *prolonged* survival in liver cancer (~2 yrs versus ~4yrs in **Fig. 3.7e,f**). Equally surprising is that no other tumor types show significance with these genes. Alternatively, survival can be assessed by comparing expression in liver tumors from many different patients (with no comparison to adjacent tissue), and this analysis shows prolonged survival for high levels of most (80%) of the contractile gene set including *ACTA2*, whereas no significance was found for survival with high levels of most (89%) of the synthetic gene set including *COL1A1* (**Fig.S3.7b,c,d**). The results at least underscore the contractile vs synthetic phenotypes made clear by SIML (**Fig. 3.6e**).

Fig. 7

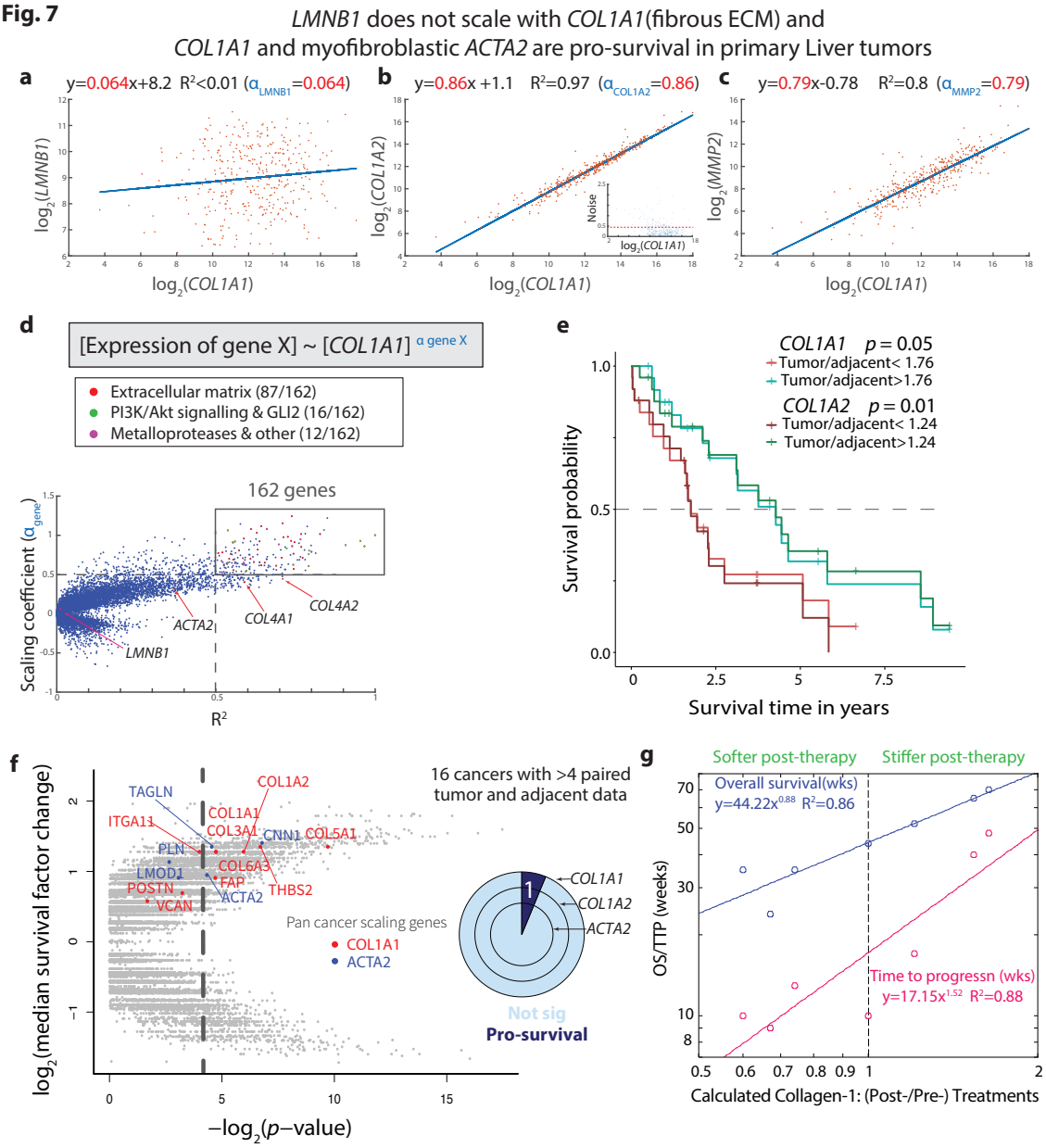


Figure 3-7*LMNB1* does not scale with *COL1A1* (fibrous ECM) and *COL1A1* and myofibroblastic *ACTA2* are pro-survival in primary Liver tumors (n=371 patients)

a-c) Plots versus *COL1A1* RNA for *LMNB1* (not correlated) and correlated ECM genes *COL1A2* and *MMP2*. The R^2 and scaling exponent (a_{gene}) are indicated. **d)** Genes that correlate with *COL1A1* RNA. For all 17,958 genes, a_{gene} versus R^2 gives a sideways-volcano plot;162 genes scale strongly a_{gene} > 0.5 and well R^2 > 0.5, and more than half relate to ECM. **e)** For *COL1A1* (light green), *COL1A2 (*dark green*)* RNA levels (tumor/adjacent) that exceed the median levels for all patients (n=50 patients), the median time for survival is significantly longer by 2-3 yrs. **f)** For all 17,958 genes, KM analyses are summarized by the fold-change in median survival plotted against the p-value, showing that many of the genes which scale with *COL1A1* or with *ACTA2* predict prolonged survival when the expression ratio (tumor/adjacent) exceeds the median. *Inset*: Liver cancer is the only one of the 16 cancers with matched adjacent tissue data that predicts survival in

patients. **g)** Liver tumor stiffness measured before therapy and 6 wks later(Qayyum *et al.*, 2019) was converted to calculated changes in collagen-1 (Col1 ~ $E^{1.5}$) and plotted versus patient survival (n=7 patients), revealing strong power laws.

3.4 Discussion

Specific values of power law exponents are informative. For example, *TOP2A* shows a larger power law in *LMNB1* scaling than *FOXM1* (1.2 vs 1.0 per **Fig.3g inset**), and inhibition of the mutual regulator FOXM1 seems to decrease TOP2A moreso than LMNB1 (**Fig.4f,g**). As a second example, high levels of the *ACTA2* contractile gene set is pro-survival in liver cancer based on (Tumor/Adjacent) data and also for the larger patient sets of tumor-only data except for one contractile gene with the lowest power: *TAGLN* shows no significance in the tumor-only data (**Fig.S7d**). All of the contractile genes scale (strongly) with *ACTA2* exponents >1. In comparison, the *COL1A1* synthetic gene set is pro-survival only in the (Tumor/Adjacent) data and is not significant in nearly all of the larger patient sets of tumor-only data (**Fig.7e,f,S7d**); and unlike the *ACTA2* exponents, all(except *POSTN*) of the synthetic genes scale (weakly) with *COL1A1* exponents <1 (**Fig.6d**), which means *COL1A1* scales super-stoichiometrically with respect to the other genes. A reasonable analogy is that collagen-1 forms the woody bulk of a tree that the other factors assemble onto like bark or leaves on the tree. Regardless of the utility of this analogy, the contractile gene set is overall more sensitive to any changes in *ACTA2.*

In our recent experiments(Dingal *et al.*, 2015), ACTA2 protein increases dramatically with a strong, switch-like cooperativity exponent of ~3 as a mechano-repressive transcription factor exits the nucleus in mesenchymal cells that spread on stiff collagenous substrates but not on soft substrates. Importantly, *COL1A1* shows no change when we overexpressed the same transcription factor in cells on rigid plastic, which drives its nuclear accumulation and represses *ACTA2* as expected. Thus specific power law exponents help to indicate strong (*ACTA2*) or weak (*COL1A1*) regulation, extending it seems to data for patient survival. It should also be noted that treatment by radiation or pharmaceutical for each patient in TCGA with liver, lung, or breast cancer shows no impact on some these key scaling results (**Fig.S3.8**).

Pro-survival results for *high* expression of fibrotic genes in liver cancer patients is surprising but is consistent with a recent study of patients having liver stiffness measured by magnetic resonance elastography (MRE) before immunotherapy and 6 wks afterwards.(Qayyum *et al.*, 2019) Remarkably, patient survival associated with *increased* local liver stiffness (rather than initial or final stiffness). Such results for stiffness (E) are readily converted to collagen-I changes through the relation COL1A1 ~ $E^{1.5}$ from mass spectrometry of tumors and diverse normal tissues(Swift *et al.*, 2013). Power law fits for Time-to-Progression and Overall Survival (in weeks) both reveal strong scaling over a ~2^2 fold calculated range of collagen-I that increases or decreases after treatment (**Fig.7g**).

Compared to the fibrosis scaling genes, the more numerous genes in the *LMNB1* scaling gene set are cell cycle related, exhibit a broader range of scaling exponents (>1 and <1), and include exactly one transcription factor *FOXM1.* We show FOXM1 regulates *LMNB1*, and that *FOXM1* and *LMNB1* predict survival in 7-9 cancer types. Liver cancer was again appropriate to illustrate survival because Lamin-B1 is not only high in malignant hepatocytes but is even a circulating biomarker(Sun *et al.*, 2010)·(Abdelghany *et al.*, 2018). Cell culture studies show *LMNB1* depletion inhibits DNA replication(Moir, Spann, *et al.*, 2000)·(Moir, Montag-Lowy and Goldman, 1994), and *LMNB1* knockout undermines development earlier than *LMNA* knockout (Harborth *et al.*, 2001b) – all of which is consistent with cell cycle scaling with *LMNB1* but not *LMNA* (**Fig. 3.3d,g, S3.3f-h**). Such a difference is also evident in UMAP plots of scRNA-seq data (**Fig.S3.6**), which confirm actively proliferating cells colocalize high *LMNB1* with *MKI67, FOXM1* and *TOP2A*, whereas *LMNA* is more broadly expressed across cell types and more similar to *COL4's* (**Fig.S3.6b,c**). Thus, even though all solid tumors show upregulation of fibrous ECM genes, especially *COL1A1* (**Fig. 3.2a,d, Table 3.1**), such changes and their impact are essentially independent of lamin-B1.

3.4.1 Scaling is interchangable, but weak scaling is lost in noisy data

Lastly, conventional heatmap presentations can suggest associations such as between *COL1A1* and *LMNB1* (**Fig. 3.2c-f**) where none exist. However, a key strength of quantitative power law relationships is the ease of predicting new power law relations:

If $A \sim B^a$ and if $C \sim B^c$, then $A \sim C^{a/c}$.

For example, defining genes *A* = *COL5A1*, *B* = *COL1A1*, and *C* = *COL1A2* with measured power laws of $a = 0.86$, $c = 0.93$ (**Fig. 3.6a**), gives $a/c = 1.08$. Measurements agree: $COL5A1 \sim COL1A2^{1.06}$ ($R^2 = 0.96$). Poorer fits yield poorer predictions of course, but weak scaling genes are also generally problematic.

Genes scale weakly or strongly with *LMNB1* and *COL1A1* across the three cancers in sideways-volcano plots (i.e. $R^2 > 0.5$) are rare or non-existent. This is surprising given evidence of genes that scale weakly ($\alpha < 0.5$) with *COL1A1* in small datasets(Cho, Irianto and Discher, 2017), but a likely reason is illustrated by the noise in *COL1A2* scaling (**Fig. 3.7b, inset**: RMSE = 0.44) – which we apply to generic power law functions *f* that are weaker (or stronger) power laws versus *COL1A1* than the actual data. Power law fits of these equally noisy *f* yield R^2 values that decrease strongly from high R^2 with the best-fit exponents, with the same trends in *LMNB1* analyses (**Fig.S3.9**). Noisy data thus explains the lack of weak scaling genes in TCGA data, and noise in the liver cancer data likely reflects batch effects (i.e. 19 batches of patient samples analyzed over ~3 yrs in several sequencing centers). Nonetheless, the strong scaling genes and especially the 'universal' power laws supported by proteomics-rheology and initial scRNA-seq (**Fig. 3.6**) should motivate new physicochemical theories for the interacting pathways that underlie proliferation and ECM production as well as patient survival.

3.5 Methods & Materials

TCGA Analysis Gene expression (mRNA-seq), copy number variation (Gistic 2 thresholded) and phenotype data was downloaded from UCSC Xena website (https://xenabrowser.net/datapages/). Primary tumor sample can be identified by the code 01, for example, TCGA-69-7978-**01** is and TCGA-69-7978-**11** is solid tissue normal or "uninvolved" adjacent tissue from the same patient. In the Liver Cancer data set, the histological type includes 6 patients with Hepatocholangiocarcinoma (one set of normal and tumor tissue both), 3 patients

with Fibrolamellar Carcinoma and 362 Hepatocellular Carcinoma patients. 32.4% of the patients in the primary tumor Liver Cancer dataset are female. If >75% of the patients showed 0 reads for a gene, i.e. out of 371 primary tumor samples in Liver Cancer, if >278 patients showed 0 reads, that gene was removed from analysis. This reduced the Illumina HiSeq gene expression dataset from 20,530 genes to 17,958 genes. For cancer staging, TCGA annotated TNM classification system was used, which pathologically stratifies the primary tumors (T) from T1-4 indicating increasing tumor size. Amount of spread into nearby structures was chosen to characterize the primary tumor size phenotype T1 (Solitary tumor without vascular invasion), T2 (Solitary tumor with vascular invasion or multiple tumors, none > 5 cm), T3 (Multiple tumors > 5 cm; or Macrovascular invasion), T4 (Tumor(s) with direct invasion of adjacent organs or with visceral peritoneum). In phenotypic treatment information, each patient has 2 rows and 2 columns -Treatment type (Radiation/Pharmaceutical) and Treatment/Therapy (Yes/No). The Liver Cancer dataset has five kinds of entries-

1. not reported
2. both Radiation and Pharmaceutical are no (untreated)
3. Only radiation is yes,
4. only pharmaceutical is yes, and
5. both radiation and pharmaceutical are yes.

Statistical Analysis Kaplan Meier plots were created using the libraries survminer and survival in R version 3.3.3. Survival fits using survfit function were made using survival object(surv), which contained time and event (death or left the study) data. The Kaplan Meier plot and p-value were obtained using the ggsurvplot function. Box-plots were made using the function boxplot in R. One-way ANOVA test (which tests significant differences in the averages of multiple data groups) was conducted using aov function in R. Linear fit models for log values of gene expression were made using the fitlm function in MATLAB 2019b using the robust regression model that iteratively reweights least squares to reduce the effect of outliers, making it less sensitive than ordinary least squares to large changes in small parts of the data. For example, sideways-volcano plots made with robust regression model for *LMNB1* in primary Liver tumors give 168 genes with high β_{gene} and R^2, but with ordinary linear regression, there are 159 genes.

R^2 measures the improvement in prediction of the variable 'y' (eg, $\log_2$(mRNA expression of *TOP2A*)) using the regression model *f*, based on predictor x (eg, $\log_2$(mRNA expressionof *LMNB1*)), compared to the 'mean model' (**Eq 3**) (i.e. does the variation in 'x' predict variation in 'y' better than just using the mean value of 'y' as the predictor).

$$R^2 = 1 - \frac{sum\ of\ squared\ errors\ of\ regression\ model}{sum\ of\ squared\ errors\ of\ baseline\ model(y=y_{mean})} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - y_{mean})^2} \qquad \textbf{Eq.3}$$

Human liver collection

Human liver samples were collected in strict accordance with the protocols approved by the Institutional Review Board of the University of Pennsylvania and the Philadelphia Veterans Administration Medical Center (protocol number 820896). De-identified samples for this research study were obtained from male and female patients aged 32 to 68 years old who underwent liver resection or transplantation. Liver tissue was initially stored in cold Hank's Balanced Salt Solution (HBSS, Gibco), then flash frozen until analysis by mass spectrometry.

Mass spectrometry (LC-MS/MS) of Liver tumors and adjacent uninvolved tissues Tissues were lysed using NUPAGE LDS buffer. The lysates were loaded onto SDS-PAGE gels (4-12%, Bis-Tris, Invitogen) and gel electrophoresis was run for 20 min at 160 V. ~1 mm^3 gel sections were carefully excised from SDS–PAGE gels and were washed in 50% 0.2 M ammonium bicarbonate (AB), 50% acetonitrile (ACN) solution for 30 min at 37°C. The washed slices were lyophilized for >15 min, incubated with a reducing agent (20 mM TCEP in 25 mM AB solution), and alkylated (40 mM iodoacetamide (IAM) in 25 mM AB solution). The gel sections were lyophilized again before in-gel trypsinization (20 mg/mL sequencing grade modified trypsin, Promega) overnight at 37°C

with gentle shaking. The resulting tryptic peptides were extracted by adding 50% digest dilution buffer (60 mM AB solution with 3% formic acid) and injected into a high-pressure liquid chromatography (HPLC) system coupled to a hybrid LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) via a nano-electrospray ion source.

Raw data from each MS sample was processed using MaxQuant (version 1.5.3.8, Max Planck Institute of Biochemistry). MaxQuant's built-in Label-Free Quantification (LFQ) algorithm was employed with full tryptic digestion and up to 2 missed cleavage sites. The software's decoy search mode was set as 'revert' and a MS/MS tolerance limit of 20 ppm was used, along with a false discovery rate (FDR) of 1%. The minimum number of amino acid residues per tryptic peptide was set to 7, and MaxQuant's 'match between runs' feature was used for transfer of peak identifications across samples. All other parameters were run under default settings. The output tables from MaxQuant were fed into its bioinformatics suite, Perseus (version 1.5.2.4), for protein annotation and sorting.

Cell lines A549 (human lung adenocarcinoma) cell lines were obtained from ATCC (American Type Culture Collection, Manassas, VA, USA) and U2OS cell lines were obtained from the laboratory of Roger Greenberg, University of Pennsylvania. ATCC provides cell line authentication test recommendations per Tech Bulletin number 8 (TB-0111-00-02; 2010). These include: microscopy-based cell morphology check, growth curve analysis, and mycoplasma detection (by DNA staining) which were conducted on all cell lines used in these studies. All cell lines maintained the expected morphology and standard growth rates with no mycoplasma detected. U2OS and A549 cell lines were cultured in DMEM high-glucose media and Ham's F12 nutrient mixture (Gibco, Life Technologies), respectively, supplemented with 10% fetal bovine serum (FBS) and 1% penicillin and streptomycin (Sigma-Aldrich). shLMNB1 retrovirus construct was obtained from the laboratory of Prof. Robert Goldman (Northwestern University, Chicago), prepared as previous described(Shimi *et al.*, 2011). GFP-LMNB1 plasmid was obtained from the laboratory of Profs. Harald Herrmann-Lerdon and Tatjana Wedig. Cells were transfected using Lipofectamine 2000 (1:1000, Thermo fisher scientific) and incubated for 3 days.

Gene editing and live cell imaging analysis CRISPR/Cas9 and homology directed repair were used to incorporate a full length GFP tag sequence in the genomic loci for H2B at the C-terminus, using mEGFP (K206A). We used the ribonucleic protein (RNP) method with recombinant wild type S. pyogenes Cas9 protein pre-complexed with a synthetic CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) duplex. Both the crRNA and the tracrRNA were purchased from Dharmacon, while the Cas9 protein was purchased from UC Berkeley QB3 Macrolab. The sequence for the crRNA is 5' ACTCACTGTTTACTTAGCGC 3', and for the tracrRNA is 5' GTCGCAGTCGCCATGGCGGG 3'.

The crRNA:tracrRNA duplex was prepared my adding crRNA (20 uM) and tracrRNA (20 uM) at a 1:1 ratio for a final concentration of 10 uM. The mixture was heated at 95°C for 5 min and then allowed to cool at room temperature for 2 hrs. After 2 hrs elapsed, 700,000 A549 RFP-LMNB1 cells were prepared and suspended in 200 uL of Ham's F12 media supplemented only with 10% FBS. Cells were placed in a 0.4 cm Gene Pulser electroporation cuvette (Bio-Rad). 10 uL of 10 uM crRNA:tracrRNA duplex, 10 uL of 10 uM Cas9 protein, and to the cell suspension was also added 8 ug of donor plasmid (AICSDP-52: HIST1H2BJ-mEGFP is Addgene plasmid # 109121). Cells were then electroporated using a Gene Pulser Xcell electroporation system at 160 V for 30 ms. The RFP-LMNB1 expressing A549 cells were as described elsewhere, with RFP tag insertion done using zinc finger nuclease at the C-terminus. The RFP tag is a monomeric TagRFP from Evrogen.

Successfully edited cells were enriched using fluorescence-activated cell sorting (FACS) with a BD FACSJazz (BD Biosciences). Cells were enriched three times using FACS. Single-cell clones were then prepared by plating 100 cells into a 15-cm plate (Corning) and isolated using cloning rings (Corning). To confirm GFP tagging of H2B by PCR, the forward primer sequence is 5' ATGCCAGAGCCAGCGAA 3', and the reverse primer sequence is 5' GAGAGTTTGCAACCAACTCACT 3'. Successful H2B integration, then a PCR product equivalent

to the sum of the GFP and H2B molecular weights is detected in addition to endogenous H2B. To confirm RFP insertion, PCR analysis was performed. RFP integration was confirmed using the RFP forward primer (sequence 5' AAGAGACCTACGTCGAGCAGC 3') and a LMNB1 reverse primer (sequence 5' AAAGCGGGAGCTCCTTCTCGGCAGT 3'). A PCR product was only detected for RFP-tagged cells with this primer set.

For imaging, A549 RFP-LMNB1 GFP-H2B were freshly sorted for double positive cells and seeded in a 24 well plate. Fresh media was replaced the next day, and cells were imaged at 20X magnification for 36 hrs with 20 image fields taken per well using the EVOS FL Auto imaging system (Thermo Fischer Scientific) at the CDB Microscopy Core at the Perelman School of Medicine.

FOXM1 Chip-Seq Analyses Based on (Davis *et al.*, 2018), FOXM1 contains a highly conserved DNA binding domain (DBD)(Weigel and Jäckle, 1990), which binds to DNA sequences containing the consensus sequence RAAAYA(Gormally *et al.*, 2014) or a canonical forkhead motif (RYAAAYA)(Major, Lepe and Costa, 2004; Sanders *et al.*, 2015)

LMNB1_1	299	h	HEK293T
LMNB1_1	216	K	K562

>hg38_encRegTfbsClustered_FOXM1 range=chr5:126775785-126776184 5'pad=0 3'pad=0 strand=+ repeatMasking=none

TTCCCCTTTCCGGGTTCAAGGGATTCTCCCACCTCAGCCTTCCAAGTAGCTGAGATTACAGGCACCCGCCATCATGC
CCGGCTAATTTTTGTATTTTTGTAGAGACGGGGTTTCACCATGTTGGTCAGGCTGGTCTTGAACTCCTGAGCTCAGGT
GATCCTCCCCCACTCGGCCTCCCAAAGTGTTGGGATTATAGGCGTGAGCCACCGCGCCCGCCCGGATGAATTTTCA
AATAG**AAACA**ATCCAGTTTCCCACATCACACGGAGATTGTGTGGATGGGTAATCCATGAAAAGCCTGGTGCTGGCAC
TCAGTCCCAATAAATGTTAGCTAAAAATTAGGAACTTTTCAAAGGGATGAATGTGCTTTCTAATAGAAGTAGCTGCCAG
TGACATCTTGGGA

LMNB1_2	698	h	HEK293T
LMNB1_2	1000	K	K562

>hg38_encRegTfbsClustered_FOXM1 range=chr5:126776393-126777089 5'pad=0 3'pad=0 strand=+ repeatMasking=none

CGCGGGGGGCGGAGGTGGCGCGGGCCGCGGCGCCGCACTCCCCAACATGGCCGGCACGCCCCGGCGCCTCCAC
GTGACTACCATGGGCCAATGGGGAGCAGGCAGCGCGCTTCGGCGGTTCGGGGGCCGGGCCTGTAAATTTAAAATCT
GGCCCTTGGGCAACCGTCAGCTCGGCCCTCCCGGCTGGGCGGTTCCCGGACCAGGCTCAGGGCTGTAAGGAAGGA
CGCCTGGGACTGCCCACCCGCGCGGCTCTGGGCCCAGGGCCCTCCCGCCGCCCCGCGGGGACCCCTTGGAGGGG
CCGGGCCTGAGAAAGGCGGGAAGGAGCCCGGAGGTTTCCGGGACGGATCTGCGCCACCGCCCACACTCGGCGGT
TCAGGGAAATT**CAAATA**TAAACGTCCCTAAAAGCCACACGCGAAACTCAGAGTGTCCCCGGCCACCGCGGGCGTGG
TAACCTTATCCCCCGCTCCGCTGGAGGGCTCGGGAAACGGCCTGGAGGCCGGGACGCTGTGGAATTACGTAATAGG
CACTTGAGTCCGAGGTTCCAAATGCACGGCTTGCAGGTTCCAAGTACCAGCCTTTATTTTTCTCTCCGGTGGAAAACT
GTATCAGGCAACTTCTTTCCATGAAAGTGAGATCTGTGCGTCTTGAAGCCGATAGATTATCACAGAAAAACCAGCAAC
CCTGATTTTAAATCATTAAAGCCCATCTCTTCCAAGCTTTCCGCACGAGAAAACAAGTGAGCCCTTCTCATTGGCCAG
ATTCCCTGTCAATCTCTCCGCTATGACGCCGAGTGGTGCCTTTTGAAGCCTCTCTAGTCCCGCCTCCCTAACCTGATT
GGTTTATTCAAACAAACCCCGGCCAACTCAGCCGTTCATAGGTGGATATAAAAGGCAAGCTACGATTGGTTCTTCTGG
ACGGAGACGGTGAGAGCGAGTCAGGGATTGGCTGGTCTGCTTCGGGCGGGCTAAAGGAAGGTTCAAGTGGAGCTC
TCCTAACCGACGCGCGTCTGTGGAGAAGCGGCTTGGTCGGGGGTGGTCTCGTGGGGTCCTGCCTGTTTAGTCGCTT
TCAGGGTTCTTGAGCCCCTTCACGACCGTCACCATGGAAGTGTCACCATTGCAGGTACGGCTCGCGGGGGGGACTG
GCGGTGGAGCCTCCGCGCGGCCCGGGCATCTCTCTGGCCGCCCGTGACGGGTGAAGCTCTGGGGCTGCGGTCAG
GCCGGCGACCGGCTTGGGAGCCCATATTCTCCATTCCCGGTTCCGGGGTGATCGTGGAGAGGCGGAAGCCCCTTCT
GGTGCT

ACTB	349	h	HEK293T

hg38_encRegTfbsClustered_FOXM1 range=chr7:5530068-5530347 5'pad=0 3'pad=0 strand=+ repeatMasking=none

CCAGAACGCACGCGCAGTTAGCGCCTTGAGTCCCAGCGCGCACGCGCAATTAGCGCCAATTCCCAGCGCGCACGC
AGTTAGCGCCCAAAGGACCAGCGCGCACGCGCAATGGCGCCCCAGCCCCCACCGGGCCTGGCGGGGGCTCCGCC

GCGCCCACCCTGCGATCCCCATTGGCAAGAGCCCGGCTCAGACAAAGACCCCGCCGGTTGCCCCCGCCCCGAGAG
CGGCACCCCCGGAGCGCGCCTCCCGAGCGCGGCCTCGCGCCTCCGAACTGGCGTGG

TOP2A_1	1000	h	HEK293T
TOP2A_1	1000	K	K562

>hg38_encRegTfbsClustered_FOXM1 range=chr17:40417593-40418395 5'pad=0 3'pad=0 strand=+ repeatMasking=none

GCCAGGCTGCCTGTCCAGAAAGCCGGCACTCAGTTTCCTCAGGAAAACGAAGCTAAGGCTCCCATTCCCCTCGCTAA
CAACGTCAGAACAGAGGACAGTTTTTAGATTTCAGGGATCTTAAATAGATTGGCAGTTCCTGGAGA**ATAAACA**TCCTT
TGCTTTTCTCCTGCACACTTTTGCCTCAGGCCACCCCTTCCCGCTTCCAAAGCCCATCTCTTCCAAGCTTTCCGCACG
AGA**AAACA**AGTGAGCCCTTCTCATTGGCCAGATTCCCTGTCAATCTCTCCGCTATGACGCCGAGTGGTGCCTTTTGA
AGCCTCTCTAGTCCCGCCTCCCTAACCTGATTGGTTTATTC**AAACA**AACCCCGGCCAACTCAGCCGTTCATAGGTGG
ATATAAAAGGCAAGCTACGATTGGTTCTTCTGGACGGAGACGGTGAGAGCGAGTCAGGGATTGGCTGGTCTGCTTC
GGGCGGGCTAAAGGAAGGTTCAAGTGGAGCTCTCCTAACCGACGCGCGTCTGTGGAGAAGCGGCTTGGTCGGGGG
TGGTCTCGTGGGGTCCTGCCTGTTTAGTCGCTTTCAGGGTTCTTGAGCCCCTTCACGACCGTCACCATGGAAGTGTC
ACCATTGCAGGTACGGCTCGCGGGGGGGACTGGCGGTGGAGCCTCCGCGCGCGGCCCGGGCATCTCTCTGGCCGCC
CGTGACGGGTGAAGCTCTGGGGCTGCGGTCAGGCCGGCGACCGGCTTGGGAGCCCATATTCTCCATTCCCGGTTC
CGGGGTGATCGTGGAGAGGCGGAAGCCCCTTCTGGTGCT

KIF18B	1000	h	HEK293T
KIF18B	1000	K	K562

>hg38_encRegTfbsClustered_FOXM1 range=chr17:44947512-44948112 5'pad=0 3'pad=0 strand=+ repeatMasking=none

TTGACAATCCCCTCTTGACAAACGGATGTGAATTTAGCCTACAGAAA**AAACA**GGTTT**AAACA**CCTGTTGGATGTCTGG
TTGCTGTAACTGGTGCTAAGCCTAGACCCGGGGACCACAGAGACTGGCTGTGGCCCCACGTATCCCAACCCTGCGT
TGGCTTCTGCCCGCTACTCTCGTGTGCCCAACCCCGGAGACACCGCCTGGGAAACTGCAGCCGACAGGAAGAGGAA
GTGAGCAAGCCCACTGTGAACGGCTCTCGCAACACTTTGTTACTTAGG**AAACA**GCCAATCAGCGATACGGACGCTG
ACCAATAGGAAGCACGGTCCGGCCCGCCCGGCTAATTTGAACGCAAGGCGCCGGGTGGAGGTTGGCGCGGGTAGG
GGAGGGTGTGGGGCGCTCTCTCCCGGTGTGGGTACTGCTGTCTGTGGTGTGGCTGTGGGACCCGTGAGCAAGCAG
CGACGCCAGCGGCGGAGAACCGACGAAAGGTAAGTCCTGGCCCACGGCGGAACCGACCGCGCGAACGCCTGCTCT
CAGGTCTCGGACGCGCAGACCCAGCTGCTCCGGGCGGGAGGCCTGTGCCAGGCCTCTTTTGTGCTGGACG

KIF20A	1000	h	HEK293T
KIF20A	1000	K	K562

>hg38_encRegTfbsClustered_FOXM1 range=chr5:138178620-138179373 5'pad=0 3'pad=0 strand=+ repeatMasking=none

CCCCGAAGTCTCCAACCCTGAGGAGAGACGGTCTGGAAGGGCCGAAACCCTGGAATTCTGGG**AAATA**GTGTACCG
GCGTTGGTTCGCGGTGCCTCCTGGGAAATGTCGTTTTTCCCCTTAAGACAAAGCAAGCACCCTAAACCAGTTACCCT
GTGCACTCCTGTTAAGATTGTTGCTAAGGAAGGACAGGAGTTGGCTGCTGAAGCCTCAAGATTTCCTTTAGGCTCTTA
GGTAAGAAATGTCTAAGGTTCAAGGAAAAAGGTTAAGTTGGAAGAATCCCAGGCA**AAATA**AGTGCGAATCCACGACA
GTTGGTAACCCGGACCCACATTAGAACTCAGAGGTCAAGCAGAAGCGAACGACTGGAATTCCAGTCAGGCCCGCCC
CCTTTCCTTACGCGGATTGGTAGCTGCAGGCTTCCCTATCTGATTGGCCGAACGAACGCAGCGCGTAATTTA**AAATA**
TTGTATCTGTAACAAAGCTGCACCTCGTGGGCGGAGTTGTGCTCTGCGGCTGCGAAAGTCCAGCTTCGGCGACTAG
GTGTGAGTAAGCCAGTATCCCAGGAGGAGCAAGTGGCACGTCTTCGGGTGAGTGTGCGGCTGTGCTGGAGCCCGG
GTTACCAGCTCTTGCCCGCGCGGAATAGGCTTTAAGAAATCGGAGCTGCAGCCATCTGTTTTTCTCTTAATGCTGCAT
CTCTGTTCTCACGGTGGTCCTTGGGGTGAGGTTAGAGAAGGAGCCGAGGGATCCTGACACCACCT

Lamin B1 promoter GFP reporter transfection Reporter construct was purchased from Genecopoeia (Rockwell) -LaminB1-reporter -GFP-HPRM15679-PF02

>HPRM15679 NM_005573; name=LMNB1;Entrez_ID=4001;Genome=hg18;chr5+:126139681-126140973;TSS=126140732;Upstream=1051,Downstream=241;Length=1293;

CACTCAGTCCCAATAAATGTTAGCTAAAAATTAGGAACTTTTCAAAGGGATGAATGTGCTTTCTAATAGAAGTAGCTGC
CAGTGACATCTTGGGAGCCTCTTTTTAAAAGCAAAAAATTATTTCATCTTAAATTTTCACAAGTTCAGATTTTTATGTAC
CTGATTTTAACGGAGTATG**AAACA**GCTTGCTCTCGTAGTTACTCACCGTCATATGTTAATCGTATCATTTCTGAGCTGT

TTTTTGATTTTACAAACCCTTAAATTAAGAAACGGATTAGGAACCGCCCATTAACACTGTTGGCAAATGATTTAAAATCA
GGGTTGCTGGTTTTTCTGTGATAATCTATCGGCTTCAAGACGCACAGATCTCACTTTCATGGAAAGAAGTTGCCTGAT
ACAGTTTTCCACCGGAGAGAAAAATAAAGGCTGGTACTTGGAACCTGCAAGCCGTGCATTTGGAACCTCGGACTCAA
GTGCCTATTACGTAATTCCACAGCGTCCCGGCCTCCAGGCCGTTTCCCGAGCCCTCCAGCGGAGCGGGGGATAAGG
TTACCACGCCCGCGGTGGCCGGGGACACTCTGAGTTTCGCGTGTGGCTTTTAGGGACGTTTATATTTGAATTTCCCT
GAACCGCCGAGTGTGGGCGGTGGCGCAGATCCGTCCCGGAAACCTCCGGGCTCCTTCCCGCCTTTCTCAGGCCCG
GCCCCTCCAAGGGGTCCCCGCGGGGCGGCGGGAGGGCCCTGGGCCCAGAGCCGCGCGGGTGGGCAGTCCCAGG
CGTCCTTCCTTACAGCCCTGAGCCTGGTCCGGGAACCGCCCAGCCGGGAGGGCCGAGCTGACGGTTGCCCAAGGG
CCAGATTTTAAATTTACAGGCCCGGCCCCCGAACCGCCGAAGCGCGCTGCCTGCTCCCCATTGGCCCATGGTAGTC
ACGTGGAGGCGCCGGGGCGTGCCGGCCATGTTGGGGAGTGCGGCGCCGCGGCCCGCGCCACCTCCGCCCCCCG
CGGCTTGCCTCCAGCCCGCCCCTCCCGGCCCTCCTCCCCCCGCCCGCCGCTCCGTGCAGCCTGAGAGG**AAACA**AA
GTGCTGCGAGCAGGAGACGGCGGCGGCGCGAACCCTGCTGGGCCTCCAGTCACCCTCGTCTTGCATTTTCCCGCG
TGCGTGTGTGAGTGGGTGTGTGTGTTTTCTTACAAAGGGTATTTCGCGATCGATCGATTGATTCGTAGTTCCCCCCCG
CGCGCCTTTGCCCTTTGTGCTGTAATCGAGCTCCCGCCATCCCAGGTGCTTCTCCGTTCCTCTAAAC

Cells were plated for 24 hours before transfection. Lipofectamine 2000 (Invitrogen #11668019) and the reporter construct were diluted in serum reduced Opti-MEM (Invitrogen #31985062), which was further diluted in cell media without penicillin, according to manufacturer's instructions. The cells were incubated in this solution for 24 hours before re-plating.

Immunofluorescence imaging Cells were rinsed with pre-warmed PBS on a shaker at low speed for 3 min, fixed with 4% paraformaldehyde (PFA, Fisher) for 15 min, washed 3 times with PBS and permeabilized with 0.5% Triton-X (Fisher) in PBS for 30 min. Permeabilized cells were then blocked with 5% BSA in PBS for ≥1.5 hrs. Samples were incubated overnight with primary antibodies in 0.5% BSA solution, with gentle agitation at 4 °C. Hoechst 33342 (1:1000, Thermo Fisher, #3570) was used to stain the DNA. The primary antibodies used were: LMNA (1:500, Cell Signaling Technology (CST), #4777), LMNB1 (Sigma-Aldrich ZRB1143-4X25UL), FOXM1 (1:1000, Abcam, # ab207298). Samples were washed three times in 0.1% BSA in PBS and incubated with the corresponding secondary antibodies at 1:500 dilution for 1.5 hrs at RT (Alexa Fluor 488 and 647 nm; Invitrogen). Cells on glass coverslips were mounted with mounting media (Invitrogen ProLong Gold Antifade Reagent). Images of adherent cells were taken with an Olympus IX81. All images in a given experiment were taken under the same imaging conditions and analyzed using ImageJ (NIH).

Western Blot Adherent cells were trypsinized, trypsin was neutralized with FBS followed by centrifugation to obtain a pellet of cells. This pellet was washed with PBS followed by centrifugation and then the cells were lysed with 1x NuPage LDS buffer ((Invitrogen; diluted 1:4 in 1x RIPA buffer,plus 1% protease inhibitor cocktail), and lysed by probe-sonication on ice (10 times with 3s pulses, intermediate power setting). 1% b-mercaptoethanol was added to the samples were then heated to 80 °C for 10 min and centrifuged for 10 min. Invitrogen Bolt 4%-12% Bis-Tris gels were loaded with 10 - 15 uL of lysate per lane. Gel electrophoresis was run for 10 min at 100 V and 30min at 160 V. Separated samples were then transferred to a nitrocellulose membrane using an iBlot Gel Transfer Device (Invitrogen). The membrane was blocked with 5% nonfat dry milk in TTBS buffer (Tris-buffered saline, BioRad; with 0.1% Tween-20), and incubated with primary antibodies against: LMNB1 (Sigma-Aldrich ZRB1143-4X25UL), GFP (Abcam ab290), HSP90AA1 (LifeSpan Biosciences LS-C352335), β-actin (Santa Cruz, #sc-47778) diluted at 1:500 at 4 °C overnight. After washing 3 times with TTBS, the membrane was incubated with 1:1000 diluted secondary Ab: anti-mouse or anti-rabbit (depending on primary antibody) HRP-conjugated IgG (GE Healthcare, Little Chalfont, UK), at room temperature for 1.5 hrs. The membrane was washed 3 times again with TTBS and developed with ChromoSensor (GenScript) for 3-10 min at RT. Immunoblot images were obtained using a HP Scanjet 4850. Densitometry was performed using ImageJ (NIH).

EdU labeling and staining EdU (10 μM, Abcam) was added to 24 well plates 1 hr before fixation. Cells were fixed and permeabilized as described above. After permeabilization, samples were stained with 100 mM Tris (pH 8.5) (MilliporeSigma), 1 mM CuSO4 (MilliporeSigma), 100 μM Cy5 azide dye (Cyandye), and 100 mM ascorbic acid (Millipore Sigma) for 30 min at room temperature. Samples were washed three times with 15-min washes on the shaker and then underwent immunostaining as described above.

Single cell RNA-seq and data analysis A549 cells were seeded in a 96 well-plate at a density of 150,000 cells/well and cultured for 2 days. Cells were then trypsinized and submitted to the Center for Applied Genomics (1014, Abramson Research Center, University of Pennsylvania) for single-cell RNA sequencing (with the 10X Genomics single-cell gene expression kit, a 400M read flow cell & single-end read using HiSeq from Illumina). 2,453 cells were sequenced with an average of 64,404 reads (and 3,457 genes) per cell. Sequencing saturation is only 49.4% and 97.5% of reads can be mapped to genome. Data were extracted from R using Cell Ranger R kit and aligned to genes via BioMart. Normalization to transcript length is not required because 10X Genomics only sequences the 3' end.

For all cells that express a particular integer value, for example 4 reads of *COL4A1*, we determine for *COL4A2* (gene *y*) the average number of non-integer reads = $\langle y \rangle$. Plotting these in a log-log plot (excluding $x = 0$ and $\langle y \rangle = 0$ reads), we determined the best-fit scaling exponent, and then repeated for *COL4A2* (gene *x*) and *COL4A1* (gene *y*) to obtain the inverse exponent. Error in y indicates S.E.M. COL4A1 and COL4A2 log-log slope fits have an inverse correlation and their averages converge to values close to 1.

Public GEO database code: GSE125449 was used in the Single cell sequencing analysis from hepatocellular carcinoma patients. The following patient codes were used in the analysis:

S02_P01_LCP21 (704 cells), S07_P02_LCP28 (124 cells), S10_P05_LCP23 (151 cells), S12_P07_LCP30 (805 cells), S15_P09_LCP38 (1046 cells), S16_P10_LCP18 (124 cells), S21_P13_LCP32 (132 cells)

For average single-cell exponents with respect to *LMNB1*, only genes showing $R^2 > 0.7$ were included. Identification of cell type based on sequencing data was performed as previously described(Ma *et al.*, 2019). Ma et al annotate clusters revealed by t-SNE using lineage specific marker genes for CAFs (*COL1A2, FAP, PDPN, DCN, COL3A1, COL6A1*), T-cells (*CD2, CD3E, CD3D, CD3G*), B cells (*CD79A, SLAMF7, BLNK, FCRL5*), TECs (*PECAM1, VWF, ENG, CDH5*), TAMs (*CD14, CD163, CD68, CSF1R*) and HPCs (*EPCAM, KRT19, PROM1, ALDH1A1, CD24*)

Visualization of Dimensionally reduced Single cell mRNA-seq Seurat(Butler *et al.*, 2018) was used to analyze Single cell sequencing data of A549 cells and HCC tumor cells which are mentioned above. For HCC tumor dataset, cells with >50% mitochondrial genes were removed since they were probably dead cells. Cell >6000 unique gene reads were removed as those are the droplets with doublets/multiplets of cells in each droplet. Cells are then normalized by total expression in each cell and then multiplied by 10000 and log transformed. Reads were then scaled, followed by principal component analysis of the variable features. This was followed by clustering analysis with resolution 0.1. Elbow Plot was used to determine that the first 10 principle components capture 90% of the variation. Uniform Manifold Approximation and Projection (UMAP) analysis was done to dimensionally reduce the first 10 principle components, followed by visualization. Similar analysis was done for the A549 cells.

3.6 Supplementary Information

Table 3. 1

Tumor types: 32 cancers (alphabetical)	Number of Primary Tumor samples:	Regulated or Dysregulated, Fig.1d:	Number of Tumor & Adjacen	Fig.3g: **17 + 2 cancers** *LMNB1*	Fig.6a: **17 cancers** *LMNB1*	Fig.6b: **17 cancers** *ACTA2*	Fig.6c: **17 cancers** *COL4A1*	Fig.6d: **17 cancers** *COL1A1*

	32 cancers	**17 + 2 cancers**	t pairs (>4), Figs.2,7 d-f: **16 cancers**					
Adrenocortical	79	Regulated	0	x	x			
Bile Duct	36	Regulated	9	x	x		x	x
Bladder	407	Regulated	19	x	x	x	x	x
Breast	1097	Regulated	114	x	x			x
Cervical	303	Dysregulated	3			x	x	x
Colon	286	Dysregulated	26			x	x	x
Endometrioid	176	Regulated	7	x	x	x		
Esophageal	184	Dysregulated	11			x	x	x
Glioblastoma	154	Regulated	0	x	x			
Head and Neck	520	Dysregulated	43				x	x
Kidney Chromophobe	66	Regulated	25	x	x	x		x
Kidney Clear Cell	533	Regulated	72	x	x	x	x	
Kidney Papillary Cell	290	Regulated	32	x	x			
Large B-cell Lymphoma	48	Regulated	0	o			x	x
Liver	371	Regulated	50	x	x	x	x	x
Lower Grade Glioma	516	Regulated	0	x	x		x	
Lung Adenocarcinoma	515	Regulated	58	x	x	x	x	x
Lung Squamous Cell	502	Dysregulated	51			x		x
Melanoma	104	Dysregulated	0					
Mesothelioma	87	Dysregulated	0					
Ocular Melanomas	80	Dysregulated	0					
Ovarian	304	Dysregulated	0				x	x
Pancreatic	178	Regulated	4	x	x	x	x	x
Pheochromocytoma & Paraganglioma	179	Dysregulated	0			x	x	
Prostate	497	Regulated	0	x	x	x	x	

Rectal	94	Dysregulated	6			x		x
Sarcoma	259	Regulated	0	x	x	x		
Stomach	415	Dysregulated	32			x	x	x
Testicular	150	Regulated	0	o		x		
Thymoma	120	Regulated	0	x	x		x	
Thyroid	505	Regulated	59	x	x			x
Uterine	57	Dysregulated	0					

Table 3. 2

24 genes that have the most increase and most decrease in expression in tumor tissue compared to normal from a subset of 50 patients from the TCGA Liver Cancer dataset. Genes in bold scale strongly with *LMNB1* mRNA expression.

GENE NAME	$log_2 \frac{RSEM\ in\ tumor}{RSEM\ in\ solid\ normal\ tissu}$	GENE NAME	$log_2 \frac{RSEM\ in\ tumor}{RSEM\ in\ solid\ normal\ tissu}$
UBE2C	3.870304	*CLEC4M*	-8.376186
BUB1	3.881966	*INS-IGF2*	-8.058758
SKA1	3.887232	*CLEC4G*	-8.043624
MAGEA1	3.89387	*CLEC1B*	-7.997756
HOXD9	3.902498	*MT1H*	-7.985588
CDC45	3.903606	*MARCO*	-7.80735
KIF2C	3.911922	*HAMP*	-7.699088
NCAPG	3.916724	*GDF2*	-7.637296
NEK2	3.929756	*CYP1A2*	-7.382774
TGM3	3.943794	*STAB2*	-7.273084
TOP2A	3.953428	*FCN2*	-7.172842
KIF20A	3.970512	*MT1G*	-7.143532
CDKN3	3.970604	*PZP*	-6.97113
GPC3	3.9977	*CXCL14*	-6.950726
MUC13	4.019738	*CRHBP*	-6.726032

NUF2	4.061704	*CNDP1*	-6.46678
IQGAP3	4.064878	*SFRP5*	-6.449706
BIRC5	4.078538	*SLCO1B3*	-6.420686
IGF2BP1	4.08192	*MT1M*	-6.239148
CDC20	4.098348	*MT1F*	-6.162274
MELK	4.112032	*CYP3A4*	-6.057148
HOXA13	4.16928	*C9*	-6.037976
CDC25C	4.227812	*COLEC10*	-5.874192
COL15A1	4.245218	*CLEC4M*	-8.376186

3.6.1 Gene overlap

Maximum k-Subset Intersection (MSI): Given a collection $C=\{S_1,...,S_m\}$ of m subsets over a finite set of elements $E=\{e_1,...,e_n\}$,and a positive integer k, the objective is to select exactly k subsets $S_{j1},...,S_{jk}$ whose intersection size $|S_{j1} \cap \cdots \cap S_{jk}|$ is maximum.

where $C=\{S_1,...,S_{32}\}$ are the 32 subsets of strong scaling genes for each tumor

$E=\{e_1,...,e_{867}\}$ is the set of 866 unique genes that scale with *LMNB1* across 32 tumors

k is the number of tumors out 32 being considered (17 in the example above)

In order to reduce the compute time, collection C was converted into a matrix where the rows are each of the elements in E and the columns are m subsets, the matrix is populated with 1 if the element e_i is present in j^{th} subset, else is 0.

For example, for LMNB1, there are 27 tumors that have 1 or more genes that scale strongly with *LMNB1*. There are 243 unique genes that are present >1 of the 27 tumor subsets. Hence there are 242 rows and 27 columns in the matrix m. If gene in row *i* is present in the strongly scaling gene subset of tumor *j*, $m_{ij}=1$ else it is 0.

To find which 17 tumors out of the 27 tumors to select such that there is intersection gene set is maximum, we do the following:

	Tumor 1	Tumor 2	Tumor 3	……	……	……	……	Tumor 27
Gene 1	1	0	1					0
Gene 2	0	0	1					1
…..								
…..								
Gene 242	0	1	1					0

For each of the $^{27}C_{17}$=8436285 combinations to choose 17 tumors from 27 tumors, we add the rows for which the sum=17 i.e the genes which are present in all those 17 tumors. The combination that shows the maximum sum of rows is the combination that shows the maximum overlap.

For *LMNB1*, the 17 tumors out 27 that show the maximum overlap of genes(25) are Thymoma, Liver, Breast, Lung Adenocarcinoma, Lower Grade Glioma, Sarcoma, Pancreatic, Prostate, Kidney Chromophobe, Adrenocortical, Kidney Clear Cell, Kidney Papillary Cell, Bile Duct, Glioblastoma, Thyroid, Endometroid and Bladder.

Below we list the number of tumors being chosen from 32 and the maximum number of genes that overlap in that group of tumors. It so happens that most times, tumors that show maximum overlap of genes for a particular number of tumors (n) being considered, are also present in the tumor group that shows maximum gene overlap in n+1 tumors with the addition of another tumor. Column 2 shows the tumor being added to the group. Different combinations can give the tie for the maximum overlap genes for that group number. For example, tumor groups of 3 – {Thymoma, Liver, *Breast*} and {Thymoma, Liver*, Lower Grade Glioma*} tie for the maximum overlap genes – 119. Similarly, for tumor groups of 8 – {Thymoma, Liver, Breast, Lung Adenocarcinoma, Lower Grade Glioma, Sarcoma, Pancreatic, *Prostate*} and {Thymoma, Liver, Breast, Lung Adenocarcinoma, Lower

Grade Glioma, Sarcoma, Pancreatic, *Kidney Chromophobe*} tie for the maximum overlap of genes – 72. For tumor groups of 13 – {Thymoma, Liver, Breast, Lung Adenocarcinoma, Lower Grade Glioma, Sarcoma, Pancreatic, Prostate, Kidney Chromophobe, Adrenocortical, Kindey Clear Cell, Kidney Papillary Cell, *Bile Duct*} and {Thymoma, Liver, Breast, Lung Adenocarcinoma, Lower Grade Glioma, Sarcoma, Pancreatic, Prostate, Kidney Chromophobe, Adrenocortical, Kindey Clear Cell, Kidney Papillary Cell, *Glioblastoma*} tie for the maximum overlap of genes – 50. As the number of tumor groups increases and the gene overlap number decreases, the number of combinations that can tie increases, for example, for tumor groups of 22 have 2302 combinations that tie at the maximum gene overlap number – 2. *KIF20A* scales with LMNB1 in 25 tumors showing overlap across the maximum number of tumors.

Number of tumors being considered	Tumor being added to the group to get maximum overlap	Maximum number of genes that overlap for the group of tumors
1	Thymoma	662
2	Liver	150
3	Breast/Lower Grade Glioma	118
4	Lung Adenocarcinoma	105
5	Lower Grade Glioma/Sarcoma	94
6	Sarcoma	88
7	Pancreatic	78
8	Prostate/Kidney Chromophobe	71
9	Kidney Chromophobe	66
10	Adrenocortical	63
11	Kidney Clear Cell/Bile Duct	57
12	Kidney Papillary Cell	53
13	Bile Duct/Glioblastoma	49
14	Glioblastoma	45
15	Thyroid	39

16	Endometroid	32
17	**Bladder**	**25**
18	Ocular Melanoma	18
19	Large B cell Lymphoma	8
20	46 possibilities (including testicular)	2
21	2 possibilities	2
22	2302 possibilities	1
23	300 possibilities	1
24	25 possibilities	1
25	1 possibility	1
26	1 possibility	0
27	1 possibility	0

Fig 3.6a plots the exponents of the 25 genes that show maximum overlap for tumor group of 17 - {Thymoma, Liver, Breast, Lung Adenocarcinoma, Lower Grade Glioma, Sarcoma, Pancreatic, Prostate, Kidney Chromophobe, Adrenocortical, Kindey Clear Cell, Kidney Papillary Cell, Bile Duct, Glioblastoma, Thyroid, Endometroid, Bladder}.

ACTA2 :

There are 1424 unique genes that scale strongly with *ACTA2* across 32 tumors, out of which 837 are are present in more than 1 tumor. Given the results in Fig.3g for LMNB1 and because 17 is a slight majority of the 32 total tumor types, we again sought 17 of 32 that show the maximum overlap of genes. To do this we apply the algorithm above for the $^{32}C_{17}$=565722720 combinations, we find 2 combinations which show the maximum overlap of 5 genes – {***TAGLN, CNN1, LMOD1, PLN****, MRVI1, ACTG2*} and { ***TAGLN, CNN1, LMOD1, PLN****, MYL9, KCNMB1*}.

Fig 3.6b plots the exponents of the 4 common genes- *TAGLN, CNN1, LMOD1, PLN*

The 2 combinationsc(in no particular order) of 17 cancers out of 32 that give equal maximum overlap of 6 genes:

Combination 1	Combination 2
Bile Duct	Bladder
Bladder	Cervical
Cervical	Colon
Colon	Endometrioid
Endometrioid	Esophageal
Esophageal	Kidney Chromophobe
Kidney Clear Cell	Kidney Clear Cell
Kidney Papillary Cell	Liver
Lung Adenocarcinoma	Lung Adenocarcinoma
Lung Squamous Cell	Lung Squamous Cell
Ovarian	Pancreatic
Pancreatic	Pheochromocytoma & Paraganglioma
Pheochromocytoma & Paraganglioma	Prostate
Prostate	Rectal
Rectal	Sarcoma
Sarcoma	Stomach
Stomach	Testicular

COL4A1 :

COL4A2 scales strongly with *COL4A1* across all 32 tumors. There are 994 unique genes that scale strongly with *COL4A1* across 32 tumors, out of which 460 are are present in more than 1 tumor. To find the 17 tumors out of 32 that show the maximum overlap of genes, we apply the algorithm above for the $^{32}C_{17}$=565722720 combinations, we find 21 combinations which show the maximum

overlap of 7 genes. We compiled a list of the genes and the number of combinations each of the genes appears in:

Gene Name	Number of 17 cancer combinations that gene appears in
COL4A2	21
COL15A1	21
NID2	21
LAMA4	18
GPR116	1
NID1	1
PDGFRB	1

Fig 3.6c plots the exponents of the 4 genes that appear in >18 combinations

COL1A1 :

There are 2 genes that scale strongly with *COL1A1* across all 32 tumors are *COL3A1* and *COL5A1*. There are 1090 unique genes that scale strongly with *COL1A1* across 32 tumors, out of which 558 are are present in more than 1 tumor. To find the 17 tumors out of 32 that show the maximum overlap of genes, we apply the algorithm above for the $^{32}C_{17}$=565722720 combinations, we find 11 combinations which show the maximum overlap of 10 genes. We compiled a list of the genes and the number of combinations each of the genes appears in:

Number of 17 cancer combinations	Gene Name

that gene appears in	
11	*COL1A2*
11	*COL3A1*
11	*COL5A1*
11	*FAP*
11	*POSTN*
10	*COL6A3*
9	*ITGA11*
7	*VCAN*
6	*THBS2*
3	*CTHRC1*
2	*ADAM12*
2	*COL11A1*
2	*ADAMTS2*
1	*COL12A1*

Fig 3.6d plots exponents of genes that appear in >6 combinations.

The 11 combinations of 17 cancers out of 32 that give equal maximum overlap of 10 genes:

1	2	3	4	5	6	7	8	9	10	11
Bile Duct	Bile Duct	Bile Duct	Bile Duct	Bile Duct	Bile Duct	Bile Duct	Bile Duct	Bladder	Bladder	Bladder
Bladder	Bladder	Bladder	Bladder	Bladder	Bladder	Bladder	Bladder	Breast	Breast	Breast
Breast	Breast	Breast	Breast	Breast	Breast	Breast	Breast	Cervical	Cervical	Cervical
Cervical	Cervical	Cervical	Cervical	Cervical	Colon	Colon	Colon	Colon	Colon	Colon

Colon	Colon	Colon	Colon	Colon	Esophageal	Esophageal	Esophageal	Esophageal	Esophageal	Esophageal
Esophageal	Esophageal	Esophageal	Esophageal	Esophageal	Head and Neck	Head and Neck	Head and Neck	Head and Neck	Head and Neck	Head and Neck
Head and Neck	Head and Neck	Head and Neck	Head and Neck	Head and Neck	Kidney Chromophobe	Kidney Chromophobe	Kidney Chromophobe	Kidney Chromophobe	Kidney Chromophobe	Kidney Chromophobe
Kidney Chromophobe	Kidney Chromophobe	Kidney Chromophobe	Kidney Clear Cell	Large B-cell Lymphoma	Large B-cell Lymphoma	Large B-cell Lymphoma	Large B-cell Lymphoma	Kidney Clear Cell	Large B-cell Lymphoma	Large B-cell Lymphoma
Kidney Clear Cell	Large B-cell Lymphoma	Liver	Large B-cell Lymphoma	Liver	Liver	Liver	Liver	Large B-cell Lymphoma	Liver	Lung Adenocarcinoma
Liver	Liver	Lung Adenocarcinoma	Liver	Lung Adenocarcinoma	Lung Adenocarcinoma	Lung Adenocarcinoma	Lung Adenocarcinoma	Liver	Lung Adenocarcinoma	Lung Squamous Cell
Lung Adenocarcinoma	Lung Adenocarcinoma	Lung Squamous Cell	Lung Adenocarcinoma	Lung Squamous Cell	Lung Squamous Cell	Lung Squamous Cell	Lung Squamous Cell	Lung Adenocarcinoma	Lung Squamous Cell	Mesothelioma
Lung Squamous Cell	Lung Squamous Cell	Ovarian	Lung Squamous Cell	Mesothelioma	Mesothelioma	Ovarian	Mesothelioma	Lung Squamous Cell	Mesothelioma	Ovarian
Ovarian	Ovarian	Pancreatic	Ovarian	Ovarian	Ovarian	Pancreatic	Ovarian	Ovarian	Ovarian	Pancreatic
Pancreatic	Pancreatic	Prostate	Pancreatic	Pancreatic	Pancreatic	Pheochromocytoma & Paraganglioma	Pancreatic	Pancreatic	Pancreatic	Pheochromocytoma & Paraganglioma
Rectal	Rectal	Rectal	Rectal	Rectal	Rectal	Rectal	Rectal	Rectal	Rectal	Rectal
Stomach	Stomach	Stomach	Stomach	Stomach	Stomach	Stomach	Stomach	Stomach	Stomach	Stomach

Thyroid	Thyroid	Thyroid	Thyroid	Thyroid	Thyroid	Thyroid	Thyroid	Thyroid	Thyroid	Testicular

3.6.2 Population Scaling from Single Cell Scaling

Consider quantities *A* and *B* that scale with the age (time since last division) *a* of cells. Then the population values of *A* and *B* are given by the averages *A(a)* and *B(a)* over the distribution $\rho(a)$ of ages a in the population. In a growing population with the doubling time τ, the age distribution is approximately given by (see derivation below)

$$\rho(a) = \begin{cases} \frac{2\ ln(2)}{\tau} 2^{-a/\tau} & a < \tau \\ 0 & a > \tau \end{cases}$$

For $A(a) \propto a^{\alpha}$ and $B(a) \propto a^{\beta}$, we have

$$A = \int_0^{\infty} \rho(a) A(a)\, da$$

$$\propto \int_0^{\tau} \frac{2ln(2)}{\tau} 2^{-\frac{x}{a/\tau}} a^{\alpha} da$$

$$= \left(2ln(2) \int_0^1 2^{-x} x^{\alpha} dx \right) \tau^{\alpha} \ \propto \ \tau^{\alpha}$$

and similarly, $B \propto \tau^{\beta}$. If *A* and *B* are sampled across cancer patients with some variability in the doubling time τ across the patients, *B* would scale with *A* as $B \propto A^{\beta/\alpha}$. While this argument only includes the variability in population doubling time τ across population samples, the resulting scaling is robust with respect to small heterogeneities in cell cycle times within each populations (see below for derivation).

Age Distribution derivation

In a growing population with the doubling time τ (approximately equal to the cell cycle time), the number of cells are given approximately given by

$$N(t) \propto 2^{t/\tau}$$

The number of cells of age *a* at time *t* is given by the number of cells of age 0 at time $t - a$.

Therefore, the distribution of ages $\rho(a)$ defined by the number of cells of age *a* divided by total number of cells is given by

$$\rho(a) \propto \frac{2^{(t-a)/\tau}}{2^{t/\tau}} = 2^{-a/\tau}$$

This argument holds for $a < \tau$ but at age τ, cells divide and therefore $\rho(a) = 0$ for $a > \tau$. The proportionality constant can be determined from the normalization

$$\int_0^{\tau} 2^{-a/\tau} da = \frac{\tau}{2ln(2)}$$

The age distribution is therefore given by

$$\rho(a) = \begin{cases} \frac{2\ ln(2)}{\tau} 2^{-a/\tau} & a < \tau \\ 0 & a > \tau \end{cases}$$

3.6.2.1 Averaging over Populations with Heterogeneous Cell-Cycle Times

We can extend these results by considering a population of cells with cell-cycle times τ_c drawn from a distribution $P(\tau_c)$. In this case the population doubling time τ_p would satisfy the following integral equation

$$\int_0^{\infty} 2^{-\frac{\tau_c}{\tau_p}} P(\tau_c) d\tau_c = \frac{1}{2}$$

and the age distribution is given by

$$\rho(a) = \frac{2 \ln(2)}{\tau_p} 2^{-a/\tau_p} \left(1 - \int_0^{a} P(\tau_c) d\tau_c \right).$$

The derivations of these relationships are cited here(POWELL, 1956; Jafarpour *et al.*, 2018). The logic behind the derivation of the age distribution is very similar to the previous case, except now, instead of cutting the distributions abruptly at some cell-cycle time, it is multiplied by the probability that a cell is not divided until age a, which reflects the fact that different cells can divide at different ages.

Now following the previous section, we can find the population average value of A as following

$$A = \int_0^\infty \rho(a)A(a)\,da$$

$$\propto \int_0^\infty \frac{2ln(2)}{\tau_p} 2^{-\widetilde{a/\tau_p}} \left(1 - \int_0^a P(\tau_c)d\tau_c\right) a^\alpha da$$

$$= \left(2ln(2) \int_0^\infty 2^{-x} x^\alpha \left(1 - \int_0^x p(u)du\right) dx\right) {\tau_p}^\alpha$$

Where $p(u) = \tau_p\, P(\tau_c)$ is the distribution of cell cycle times rescaled by the population doubling time. If this distribution is similar across different samples (i.e. the relative variability in doubling times are the same across samples), we are essentially done: we have shown that $A \propto {\tau_p}^\alpha$ and similarly, $B \propto {\tau_p}^\beta$, and therefore, B would scale with A as $B \propto A^{\beta/\alpha}$.

What if $p(u)$ is different across samples? Let us approximate the cell cycle time distribution by a gamma distribution of mean μ and standard deviation σ. Note that narrow gamma distributions are approximately Gaussian by central limit theorem, but they have the advantage of having their support on positive reals. For gamma distributed $p(u)$, one can show that A is approximately proportional to

$$A \propto \left(1 + C_\alpha \left(\frac{\sigma}{\mu}\right)^2\right) {\tau_p}^\alpha ,$$

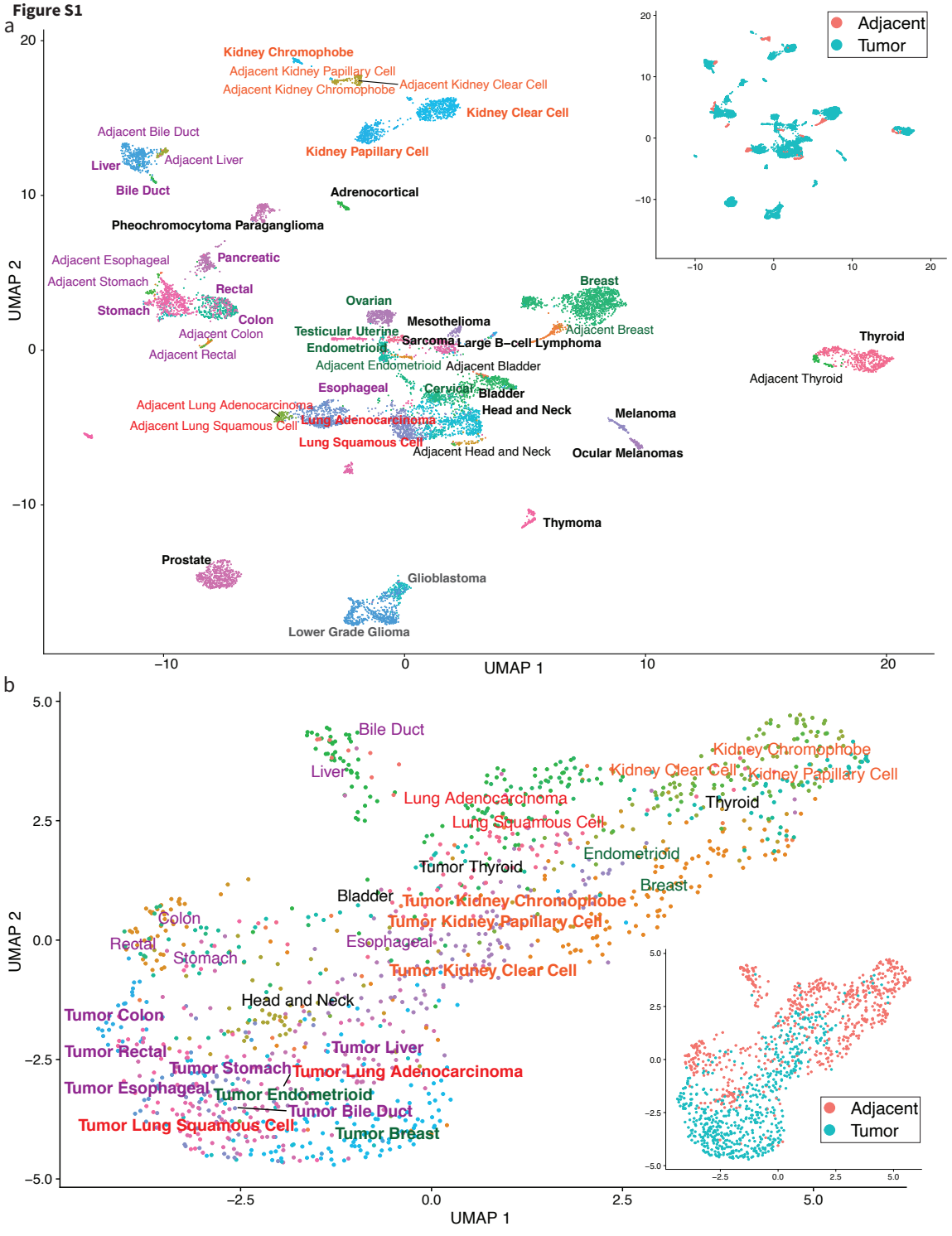
where C_α is an α-dependent constant of order 1 ($C_1 \approx 0.7$ and $C_2 \approx 1$). As long as the variance of $(\sigma/\mu)^2$ across the samples is significantly smaller than the relative variability in population doubling times, the derived scaling of $B \propto A^{\beta/\alpha}$ holds. Note that this analysis is done for narrow distributions of cell-cycle times, which already implies $(\sigma/\mu)^2 \ll 1$ which in turn implies that $\mathrm{Var}[(\sigma/\mu)^2] \ll 1$.

3.6.3 Supplementary figures

Figure S3. 1

a) Dimensionally reducing primary tumor (32) and adjacent (16) bulk mRNA sequencing data using UMAP clusters adjacent uninvolved tissue with the corresponding tumor tissue but does not cluster based on adjacent/tumor identity(*inset)*
b) Informing UMAP with scaling information of *LMNB1* clusters the adjacent tissues away from the primary tumor tissues

Figure S1


a
Kidney Chromophobe
Adjacent Kidney Papillary Cell
Adjacent Kidney Chromophobe
Adjacent Kidney Clear Cell
Kidney Clear Cell
Adjacent Bile Duct
Kidney Papillary Cell
Liver
Adjacent Liver
Bile Duct
Adrenocortical
Pheochromocytoma Paraganglioma
Adjacent Esophageal
Pancreatic
Adjacent Stomach
Rectal
Stomach
Colon
Adjacent Colon
Adjacent Rectal
Ovarian
Testicular Uterine
Endometrioid
Mesothelioma
Sarcoma
Large B−cell Lymphoma
Adjacent Endometrioid
Adjacent Bladder
Breast
Adjacent Breast
Thyroid
Adjacent Thyroid
Esophageal
Cervical
Bladder
Head and Neck
Adjacent Lung Adenocarcinoma
Adjacent Lung Squamous Cell
Lung Adenocarcinoma
Lung Squamous Cell
Adjacent Head and Neck
Melanoma
Ocular Melanomas
Thymoma
Prostate
Glioblastoma
Lower Grade Glioma
UMAP 1
UMAP 2
Adjacent
Tumor
b
Bile Duct
Liver
Kidney Chromophobe
Kidney Clear Cell
Kidney Papillary Cell
Lung Adenocarcinoma
Thyroid
Lung Squamous Cell
Endometrioid
Tumor Thyroid
Breast
Bladder
Tumor Kidney Chromophobe
Tumor Kidney Papillary Cell
Colon
Rectal
Stomach
Esophageal
Tumor Kidney Clear Cell
Head and Neck
Tumor Colon
Tumor Rectal
Tumor Liver
Tumor Stomach
Tumor Lung Adenocarcinoma
Tumor Esophageal
Tumor Endometrioid
Tumor Bile Duct
Tumor Lung Squamous Cell
Tumor Breast
UMAP 1
UMAP 2
Adjacent
Tumor

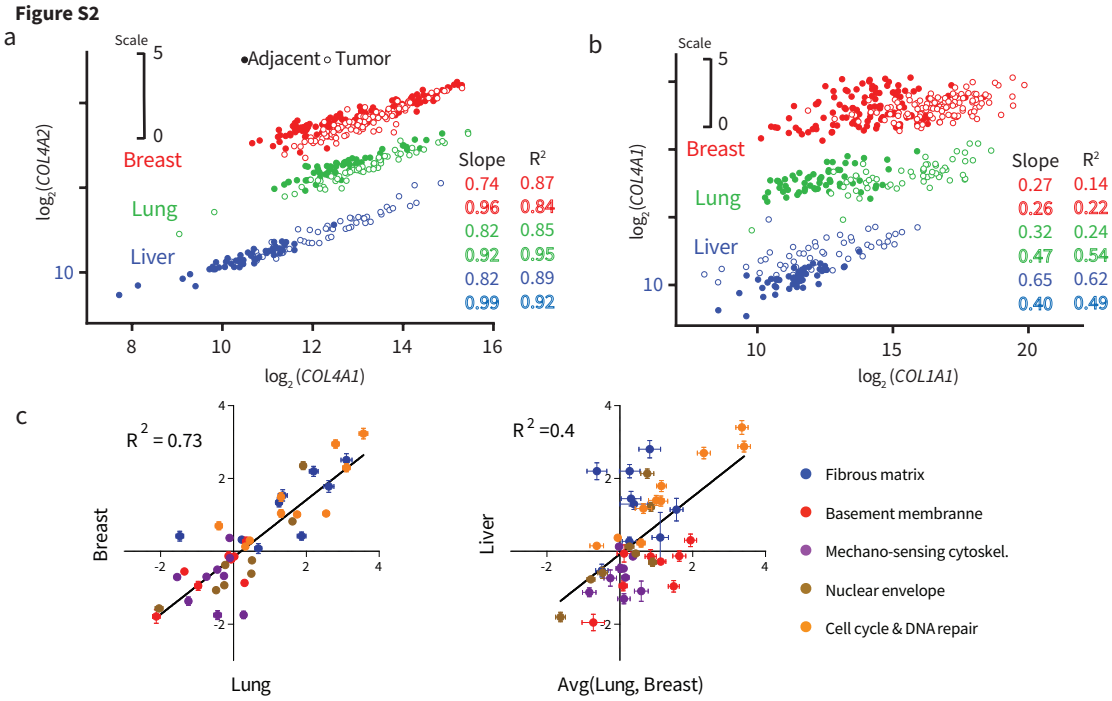


Figure S3. 2

a) Plots of $\log_2$(*COL4A1*) vs $\log_2$(*COL4A2*) should show an excellent correlation as subunits of the type IV collagen, which are also present very close to each other on the same chromosome and have a shared promotor. Tumor (open circle) and normal (closed circle) tissue data points are plotted individually for Liver (blue), Lung Adenocarcinoma (green) and Breast (red) Cancer. All three cancer datasets show excellent R^2 (0.84 to 0.95). Tumor data points for only Liver cancer show higher *COL4A1* and *COL4A2* values. (Note that for better visibility, all Lung Adenocarcinoma data points were moved up by 3 units and Breast Cancer by 6 units on the y-axis.)
b) Plots of $\log_2$(*COL1A1*) vs $\log_2$(*COL4A1*) show poor R^2 (0.14 to 0.62) for most tissue types implying that the basement membrane ECM is regulated differently than fibrous ECM. (Note that for better visibility, all Lung Adenocarcinoma data points were moved up by 5 units and Breast Cancer by 10 units on the y-axis.)
c) Plotting the fold change in mRNA for Lung vs Breast Cancer shows that genes show similar trends of up/downregulation (R^2 = 0.73). However, plotting the avg(Lung, Breast) vs Liver Cancer shows a weak correlation (R^2 = 0.4)

Figure S3

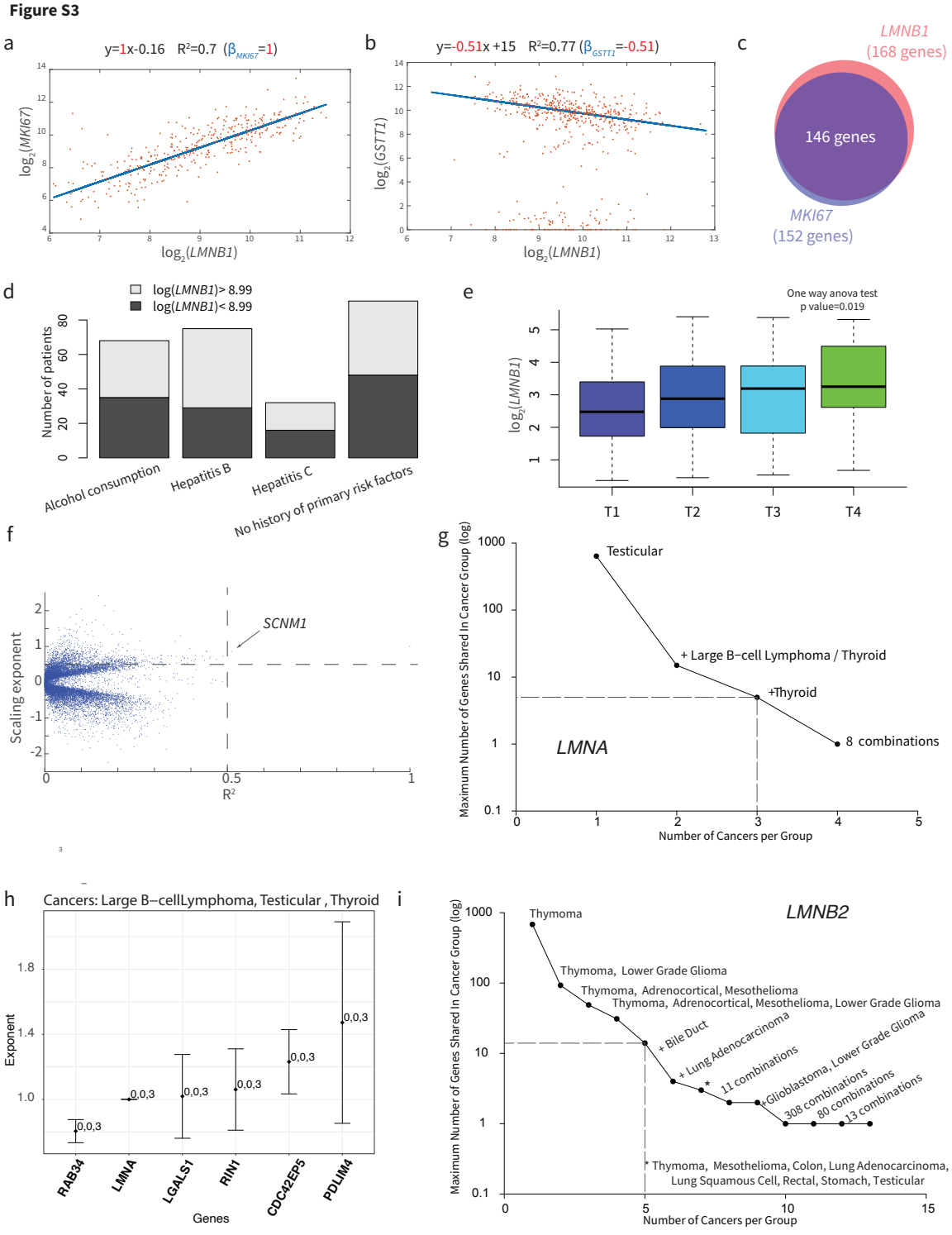


Figure S3. 3

a) Plot of $\log_2(LMNB1)$ vs $\log_2(MKI67)$ in primary Liver Cancer tumor (n=371 patients).
b) Scatterplot of the only gene that shows a negative power law relation versus *LMNB1* with a good R^2 = 0.77 in primary Lung Adenocacinoma (n=515 patients), *GSTT1*.

c) 153 genes are parsed out with slope > 0.5 and R^2 > 0.5 in the sideways-volcano plot for *MKI67*, 146 of which are identical to genes that scale with *LMNB1*.
d) Plotting the distribution of primary risk factors including alcohol consumption, hepatitis B and hepatitis C shows no difference between high and low *LMNB1* mRNA expressing patients.
e) Box and whisker plot of *LMNB1* mRNA stratified on the basis of stage of cancer progression (T1-T4). Increase in median *LMNB1* expression occurs with progression of tumors, with a significant p-value (one-way ANOVA).
f) Sideways-volcano plot for *LMNA* in primary tumors of Liver Cancer (n=371) shows only one gene with high scaling coefficient (>0.5) and R^2 > 0.5, which supports the notion that the two lamin isoforms serve distinct functions and reveals *LMNA* does not directly scale with cell proliferation genes.
g) Maximum overlap of genes drops down to 1 gene for groups of 4 tumors for LMNA strong scaling genes
h) Of the 32 tumors, the 3 tumors that show maximum overlap of 5 genes are Large B-cell lymphoma, Testicular and Thyroid

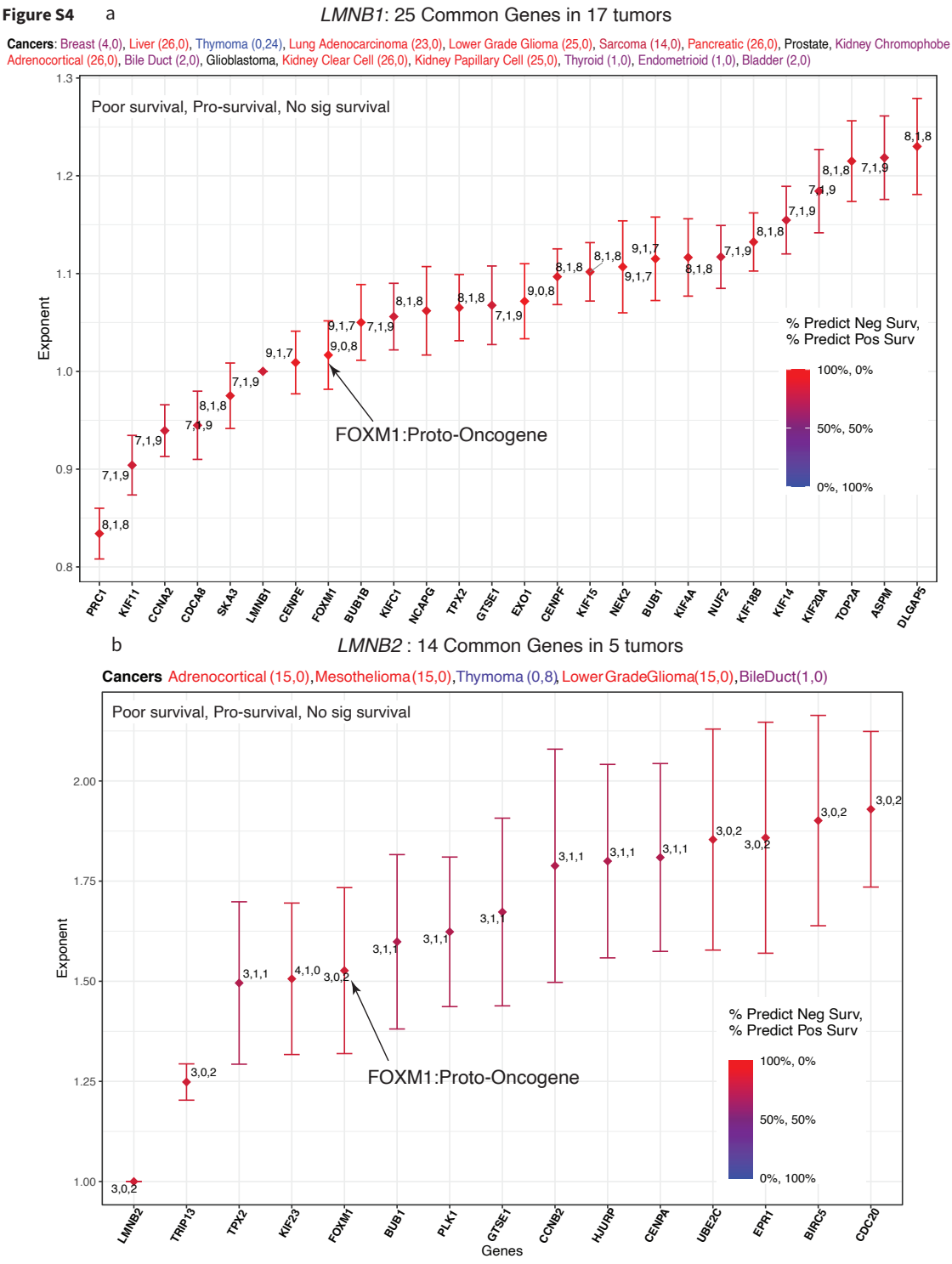


Figure S3. 4

a) Average exponents of the 25 genes that show maximum overlap in groups of 17 out of 32 tumors for *LMNB1* annotated with number of tumors where higher expression of the gene predicts survival in patients as (poor, longer or no significant trend). The 17 cancers are listed as well, indicating the

number of genes among the 25+1(*LMNB1*), higher expression for which significantly predicts (poor survival, longer survival)
b) Similarly, for LMNB2, average exponents of the 14 genes that show maximum overlap in groups of 5 out of 32 tumors

Figure S3. 5

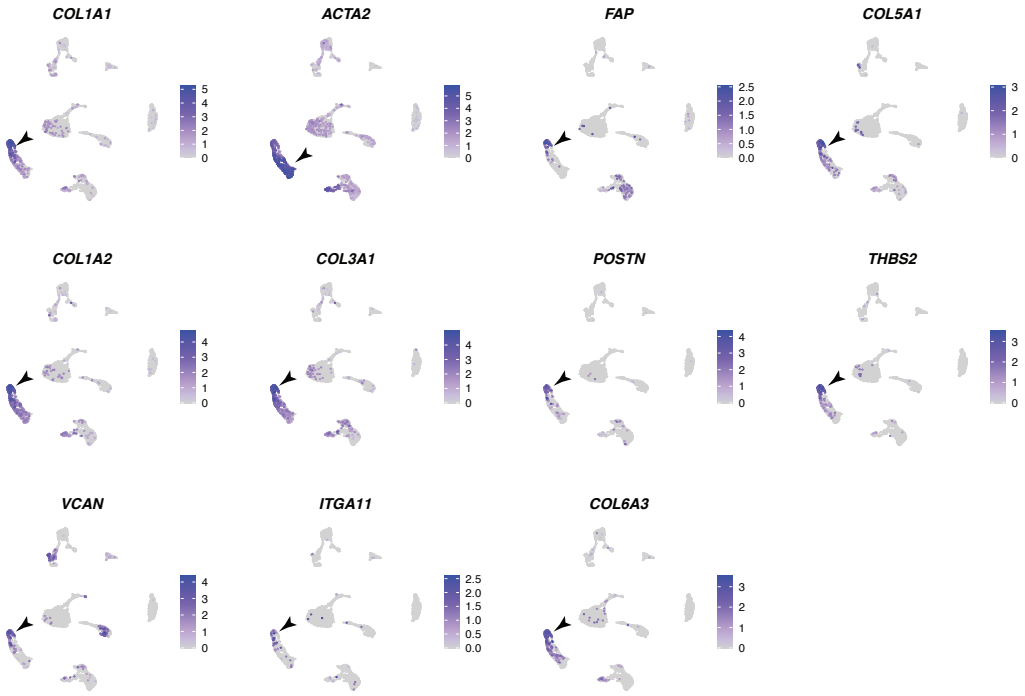
a) Cancer associated fibroblasts(CAFs) diverge as two sub-populations as synthetic (high reads of *COL1A1* scaling genes) and contractile (high reads of *ACTA2* scaling genes)

Figure S3. 6

a) Dimensionally reduced projection (UMAP) of single-cell mRNA sequencing of HCC tumors depicting clusters based on sequencing profiles labeled with the corresponding cell-type phenotype
b) UMAP from (a) labeled with expression levels of *FOXM1, TOP2A, LMNB1, LMNA, COL4A1, COL4A2, ACT2A2, COL1A1, COL1A2, MYL9, MKI67, CDK1* in each cell of the clusters
c) Dimensionally reduced projection (UMAP) of single-cell mRNA sequencing of 4 genetically diverged clones of A549 cells (P1, P2, P3, P4) labeled with expression levels of *LMNB1, FOXM1, TOP2A, COL4A1, COL4A2, LMNA*

Figure S5

Pan-cancer *COL1A1* scaling genes identify synthetic fibroblasts



Pan-cancer *ACTA2* scaling genes identify contractile fibroblasts

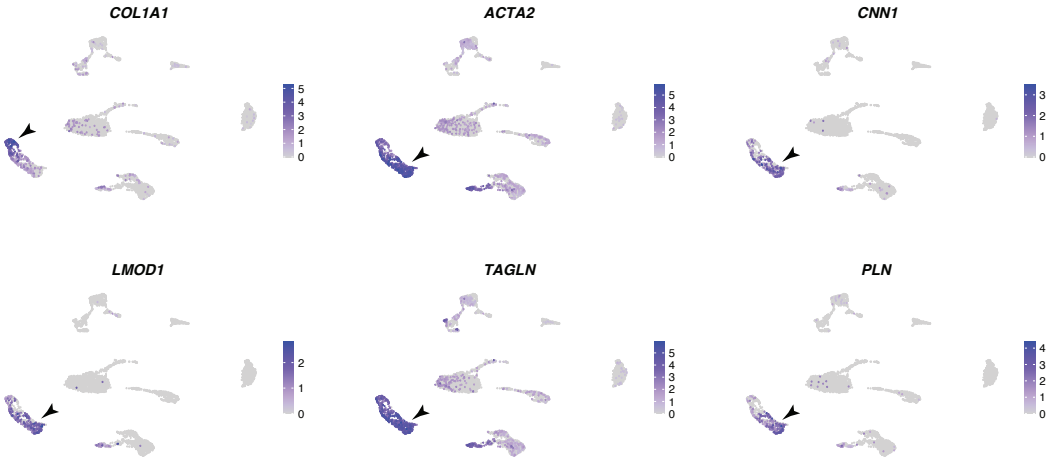


Figure S6

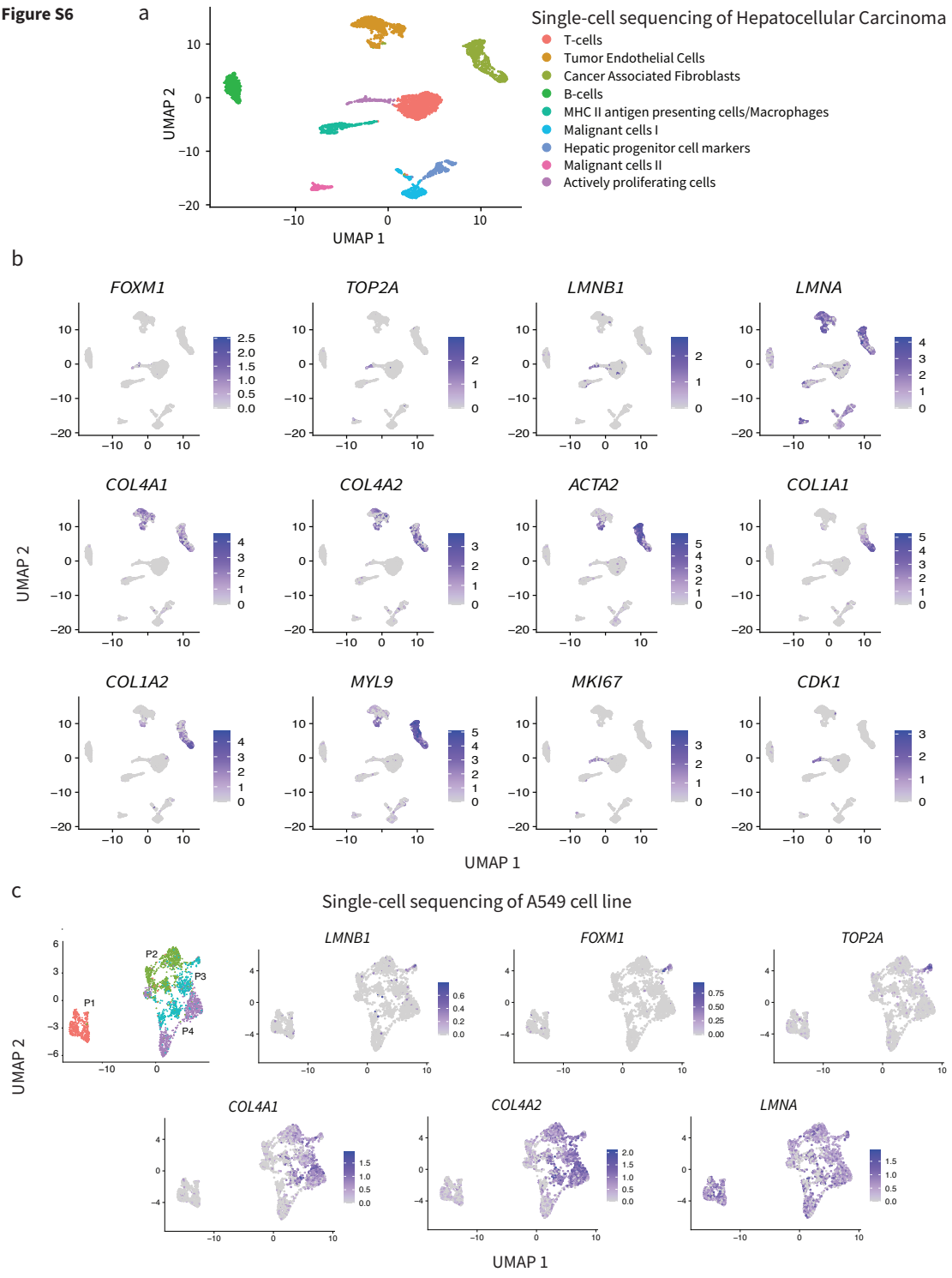

a
Single-cell sequencing of Hepatocellular Carcinoma
T-cells
Tumor Endothelial Cells
Cancer Associated Fibroblasts
B-cells
MHC II antigen presenting cells/Macrophages
Malignant cells I
Hepatic progenitor cell markers
Malignant cells II
Actively proliferating cells
UMAP 1
UMAP 2
b
FOXM1
TOP2A
LMNB1
LMNA
COL4A1
COL4A2
ACTA2
COL1A1
COL1A2
MYL9
MKI67
CDK1
UMAP 1
UMAP 2
c
Single-cell sequencing of A549 cell line
P1
P2
P3
P4
LMNB1
FOXM1
TOP2A
COL4A1
COL4A2
LMNA
UMAP 1
UMAP 2

Figure S7

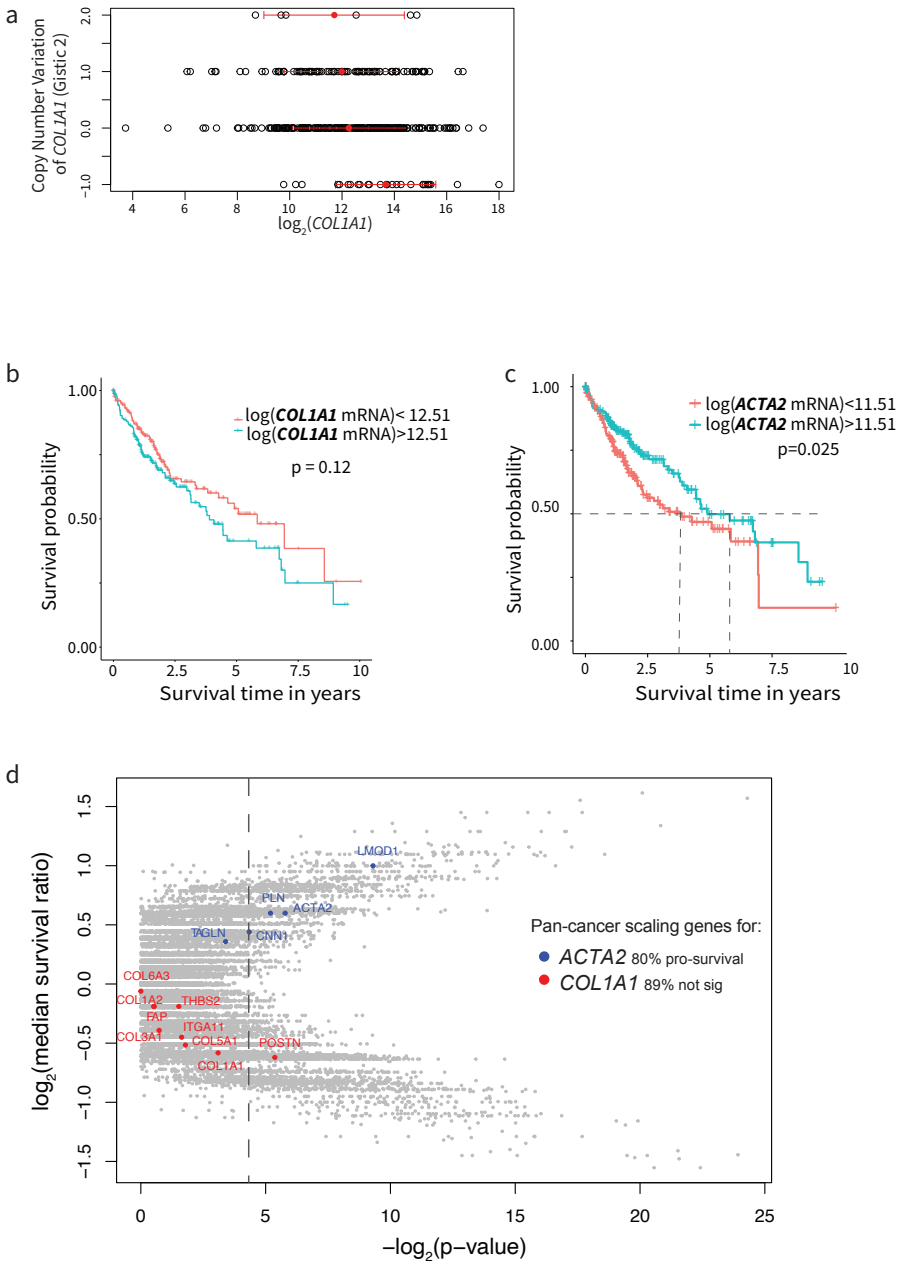


Figure S3. 7

a) *COL1A1* mRNA for primary Liver Cancer tumor (n=371) vs the corresponding threshold Gistic 2 copy number variation data for the same patient. The Gistic2 values have a thresholding of -2, -1,0,1,2, representing homozygous deletion, single copy deletion, diploid normal copy, low-level

copy number amplification, or high-level copy number amplification respectively. The red points indicate the mean mRNA expression values.
b) *COL1A1* mRNA levels alone are not a good indicator of survival probability.
c) Dividing the range of marker for activated fibroblasts- *ACTA2* mRNA levels (in primary tumor of 371 patients) into two groups, one transcribing higher *ACTA2* mRNA(greater than the median) and the other lower, and plotting a Kaplan Meier(KM) using the phenotype data available for time of death, we find that patients translating higher *ACTA2* mRNA in the tumor tissue possess a significantly lower risk of death than ones with lower *ACTA2* (median survival (number of years after which survival probability = 0.5) reduced by ~2 years).
d) For all 17,958 genes, Kaplan Meier analyses are summarized by the fold-change in median survival plotted against the p-value, yielding 3,464 genes that show significant differences, with 80% of pan cancer ACTA2 scaling genes predicting prolonged survival

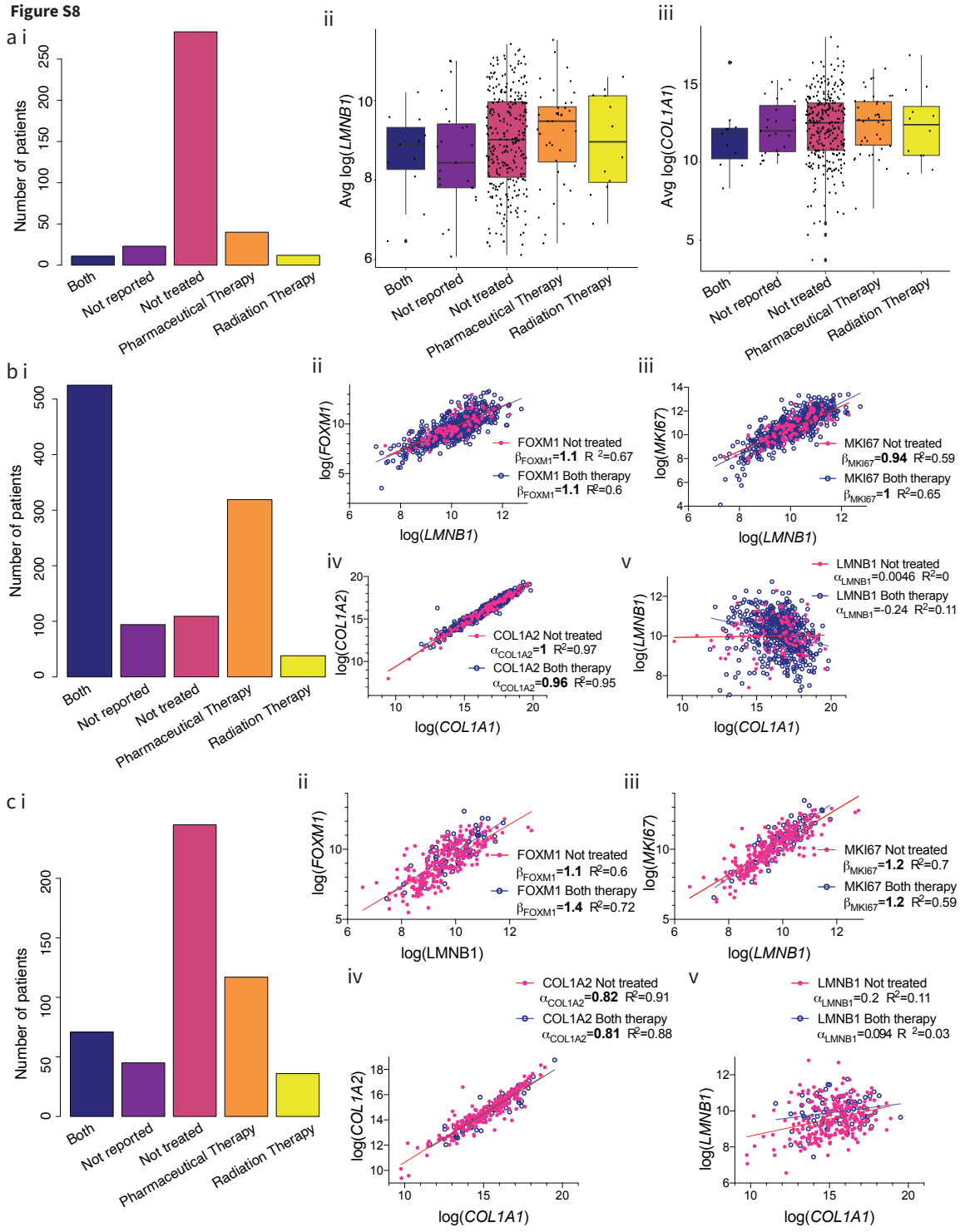


Figure S3. 8

a)i) ~75% of the primary tumor data patient in Liver cancer were not treated. Average **ii)** *LMNB1* and **iii**) *COL1A1* reads are not affected by treatment

b)i) ~48% of Breast cancer primary tumor data patients were treated with both pharmaceutical and radiation therapy. Scaling of *LMNB1* with mitotic genes **ii**) *FOXM1* & **iii**) *MKI67* and **iv**)collagen is

unaffected in patients that are treated with both radiation and pharmaceutical therapy or patient which are not treated at all. **v**) There is no correlation between *COL1A1* and *LMNB1* irrespective of treatment in Breast Cancer

c)i) Lung adenocarcinoma shows 47% of the patients are not treated, however we compare the 71 patients that are treated with both radiation and pharmaceutical therapy with them to find that scaling exponents are robust to therapy given to patients

Figure S8

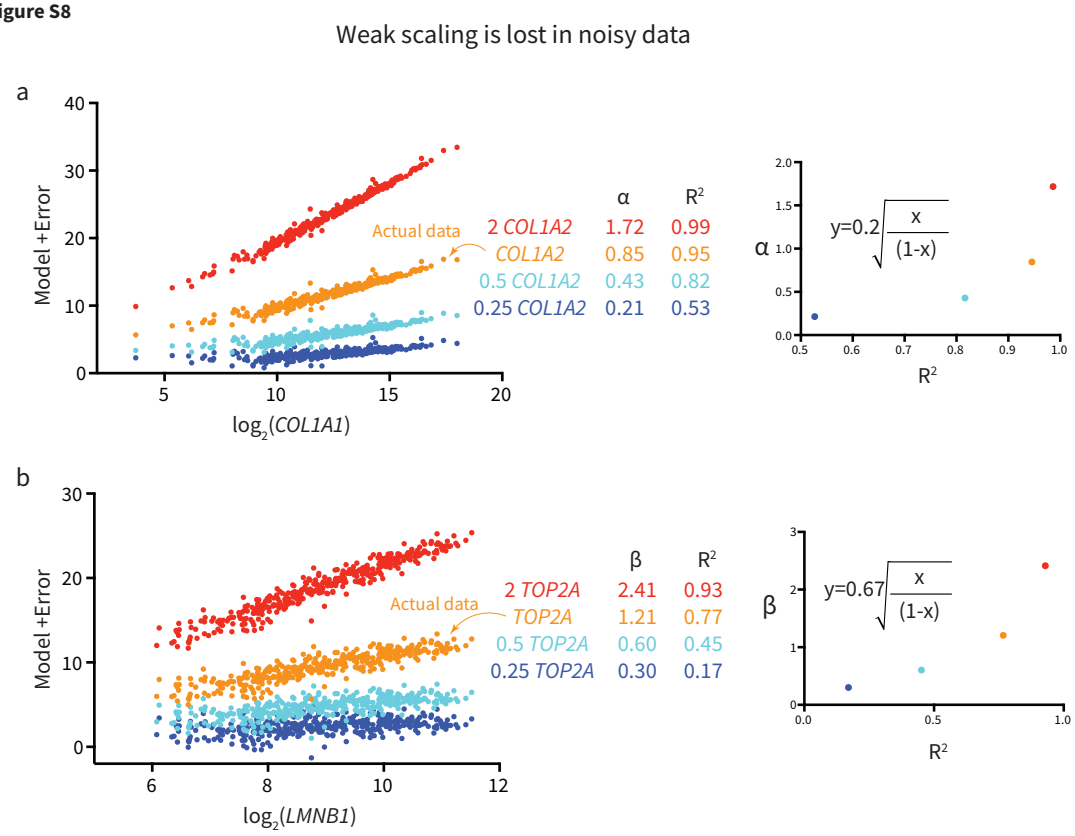


Figure S3. 9 Weak scaling can be lost in noisy data.

a) Multiplying the fit of $log_2([mRNA\ expression\ of\ COL1A1])$ vs $log_2([mRNA\ expression\ of\ COL1A2])$ (orange) by 2 (red), 0.5 (cyan) and 0.25 (blue) and adding back the residuals of the fit with a RMSE of 0.44, gives lower R^2 for lower slope. Right plot: Fitting the different slopes vs R2 for the same RMSE.

b) Similarly for model fit obtained from $log_2([mRNA\ expression\ of\ LMNB1])$ vs $log_2([mRNA\ expression\ of\ TOP2A])$ with a higher RMSE of 0.82.

Chapter 4 A study in DNA damage repair

4.1 Introduction

Single stranded DNA breaks are quite common and quickly repaired by using the other strand as a template through various repair mechanisms like Base excision repair (BER), Nucleotide excision repair (NER) and mismatch repair. However, double strand DNA breaks (DSBs) are particularly hazardous and can lead to major genomic rearrangements. ATM serine/threonine kinase is activated and recruited to DSB sites and phosphorylates histone H2AX at serine 139. This γH2AX acts as a marker for DNA damage(Huang, Halicka and Darzynkiewicz, 2004). DSBs are tricky because both strands of DNA are broken, except for the existence of the sister chromosome, the information at the break is potentially lost. Reinstatement of the information depends on the accuracy of the DSB repair pathway. There are 3 major pathways to repair DSBs:

1. Homologous Recombination HR (key proteins- *BRCA1, BRCA2*): It is the most accurate form of DSB repair since it uses the sister chromatid as the template for repair. If there is a 3' overhang at the site of DNA break and the cell is in S or G2 phase of the cell cycle, the break is repaired by HR. S/G2 phase has 2-3 other sister chromatids to be used as template for HR and hence it is preferred. In G1 phase of the cell cycle, HR is inhibited by 53BP1(S and SJ, 2014).
2. Non homologous end joining c-NHEJ (key proteins- Ku80, Ku70, 53BP1): It simply detects and joins the ends of the broken DNA without any template and can sometimes lead to insertions. Hence telomere which are ends of chromosomes, if not protected by the shelterin complex (mainly TRF2), results in telomere end joining due to NHEJ (B, A and T, 1998).
3. Microhomology mediated end joining MMEJ or alt-NHEJ (key proteins Lig3, PolQ): It is highly error prone and initiates several genomic rearrangements promoted by polymerase theta (PolQ) which inhibits HR (PA *et al.*, 2015)

Nuclear membrane plays the important role of separating the nucleoplasm from the cytoplasm and maintaining the concentration of nuclear proteins like DNA repair factors. Nuclear membrane rupture is observed in cells migrating through narrow transwells or cells with low LMNA cultured on 2D glass slides (Irianto *et al.*, 2017; Xia, Ivanovska, *et al.*, 2018).This leads to sustained mis-localization of DNA repair factors including Ku80 from the nucleus to the cytoplasm. Culturing cells on stiff matrix or knocking down LMNA leads to an increase in the percentage of nuclei that are ruptured(Xia, Ivanovska, *et al.*, 2018), as well as passage through narrow pores(3um) vs larger pores (8um) leads to increase in rupture. These cells also show increase in DDR (DNA Damage Response) measured by increase in gH2AX foci as well as cell cycle arrest(Irianto *et al.*, 2017). This phenotype is also seen in cells with knockdown of repair factors (siKu80, siBRCA1), which is then rescued by simultaneous transfection of exogenous repair factors (GFP Ku80, GFP Ku70 and GFP BRCA1). On the other hand, knockdown of Ku80 in mice leads to a progeroid phenotype(Reiling *et al.*, 2014). Their data confirms that inactivation of Ku80 and DNA-PKCS causes reduced lifespan and bodyweights, which is most severe in ku80-/- mice. All mutant mice exhibited a strong increase in lymphoma incidence as well as other aging-related pathology (skin epidermal and adnexal atrophy, trabacular bone reduction, kidney tubular anisokaryosis, and cortical and medullar atrophy) and severe lymphoid depletion. Similarly, progeroid phenotype is observed with LMNA mutations. Numerous mutations in the human A-type lamin gene (*LMNA*) cause the premature aging disease, progeria (Taimen *et al.*, 2009). The Ku and TRF1 complex are a specific high-affinity interaction, as demonstrated by several in vitro methods, and exists in human cells as determined by coimmunoprecipitation experiments (Hsu *et al.*, 2000). Following targeted Ku80 knockdown, telomere defects are observed and the steady state levels of the TRF2 protein are reduced. Inhibitor studies indicate that this loss of TRF2 is mediated by the proteasome and degradation of TRF2 following Ku depletion appears to involve a decrease in chromatin binding of TRF2, suggesting that the Ku heterodimer enhances TRF2 chromatin association and that non-chromatin bound TRF2 is targeted to the proteasome. Also, chromosome end-to-end fusions are inhibited by shelterin; a multisubunit complex anchored to telomeric DNA by two Myb-containing proteins—TRF1 and TRF2 (A, 2012). Telomere fusions are executed by two independent end-

joining pathways. Classical non-homologous end joining (C-NHEJ), mediated by LIG4 and the Ku70/80 heterodimer, is primarily blocked by TRF2 (B, A and T, 1998) . Conversely, alternative NHEJ (alt-NHEJ), which is dependent on LIG3 (H *et al.*, 2005) and PARP1 (M *et al.*, 2006), is repressed in a redundant manner (R *et al.*, 2010; A and T, 2012) . Alt-NHEJ is fully unleashed after the simultaneous deletion of TRF1 andTRF2, and the creation of shelterin free telomeres in cells deficient for Ku70 and Ku80 (also known as Xrcc6 and Xrcc5, respectively)(A and T, 2012). This error-prone end-joining pathway mediates fusion of naturally eroded telomeres1, joining of switch regions during class-switch recombination, and formation of chromosomal translocations in mouse cells (Simsek and Jasin, 2010; Simsek *et al.*, 2011). Despite its central role in NHEJ, reducing the levels of Ku80 in human tumor cell lines leads to telomere shortening, telomere fusion and apoptosis as the primary phenotype (I, P and MA, 2004; K *et al.*, 2004; G *et al.*, 2007). It has been reported that telomere dysfunction in immortal human cells resulting from complete Ku depletion is accompanied by telomere uncapping and the appearance of telomeric circles (Wang, Ghosh and Hendrickson, 2009). Thinking about the overarching concepts of an aging phenotype displayed by both knockdown of Ku80 as well as LMNA mutations, also NHEJ and MMEJ DNA repair pathways being implicated in telomere fusions, in this chapter we study the relationship between rupture of nuclear membrane and subsequent mis-localization of repair factors with attrition of telomeres, cell cycle arrest and aging.

Oxidative stress induced DNA damage is a major concern in heart because it is a very energy intensive organ(Lin *et al.*, 2006; Mondal *et al.*, 2013). Mitochondria produce harmful reactive oxidative species (ROS) as a byproduct of their metabolic process and hence play a regulatory role in the cell apoptotic process as well. Previous studies in cancer cells show partial rescue of DNA damage though antioxidants(Xia *et al.*, 2019). Chapter 2 shows inhibtion of acto-myosin stress can rescue DNA damage in embryonic hearts. We now want to study if DNA damage rescue through acto-myosin stress inhibition acts independently from rescue through antioxidants. This is to parse out if reduction in DNA damage due to blebbistatin treatment can happen due to reduced mitochondrial stress as well as reduced mis localization of DNA repair factors studied in chapter 2.

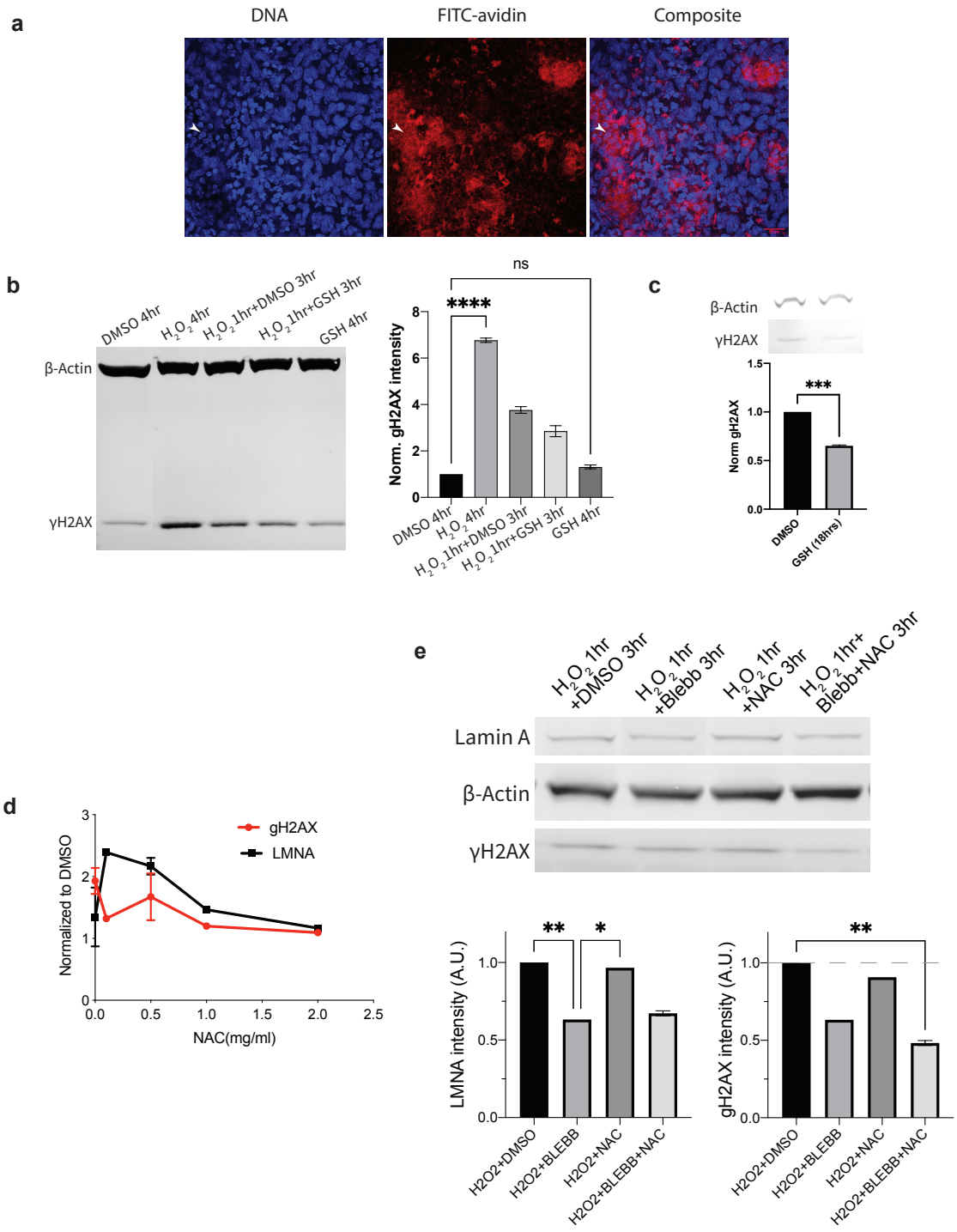
4.2 Results

4.2.1 Oxidative stress as a source of DNA Damage in hearts and interplay with acto-myosin stress

Reactive oxidative species (ROS) cause oxidative DNA damage by oxidizing DNA base pair guanine (G) to 8-oxo-G. It is repaired by the Base excision repair (BER) pathway. Histone H2AX has also been shown to by phosphorylated through the BER pathway and hence can be used as marker for oxidative DNA damage (Marti *et al.*, 2006). FITC-avidin binds to 8-oxo-G which resembles the structure of biotin and can be used to measure DNA damage specific to ROS. However in chicken embryonic hearts, FITC has a diffuse cytoplasmic signal showing high intensity around small nuclei that might be undergoing apoptosis (**Fig 4.1a**). Ratio of reduced glutathione (GSH) and oxidized glutathione disulphide (GSSG) is an important indicator of the oxidative state of a cell (Puente *et al.*, 2014). We find that gH2AX shows a significant increase when the hearts are treated with hydrogen peroxide (H_2O_2) confirming that gH2AX is good marker for oxidative DNA damage (**Fig 4.1b**). We use GSH as an anti-oxidant to study rescue of oxidative damage in embryonic hearts. Glutathione reduced ethyl ester (GSH-MEE) is not able to rescue basal oxidative DNA Damage after 4 hours, but we find significant decrease DNA damage in hearts treated with GSH for 18 hours compared to control hearts kept in the same incubator after extraction from chicken egg (**Fig 4.1c**).

We also study the use of physiologically relevant N-acetyl cysteine (NAC) in rescue for oxidative stress. First we study a dose dependent response to DNA damage with hearts treated with NAC at different concentrations for 3 hrs. We find that 2mg/ml is the optimum concentration (**Fig 4.1d**) and at higher concentration of 5mg/ml there is instead an increase in DNA damage. Using 1 hr H_2O_2 treatment an external source of oxidative stress, followed by 3 hrs of treatment by blebbistatin or anti-oxidant or both, we find that acto-myosin stress inhibition and reducing oxidative stress acts in a co-operative manner to significantly reduce DNA damage. Acto-myosin stress inhibition reduces DNA damage to a greater extent than anti-oxidant rescue. This indicates that acto-myosin stress inhibition and anti-oxidants rescue DNA damage independently.

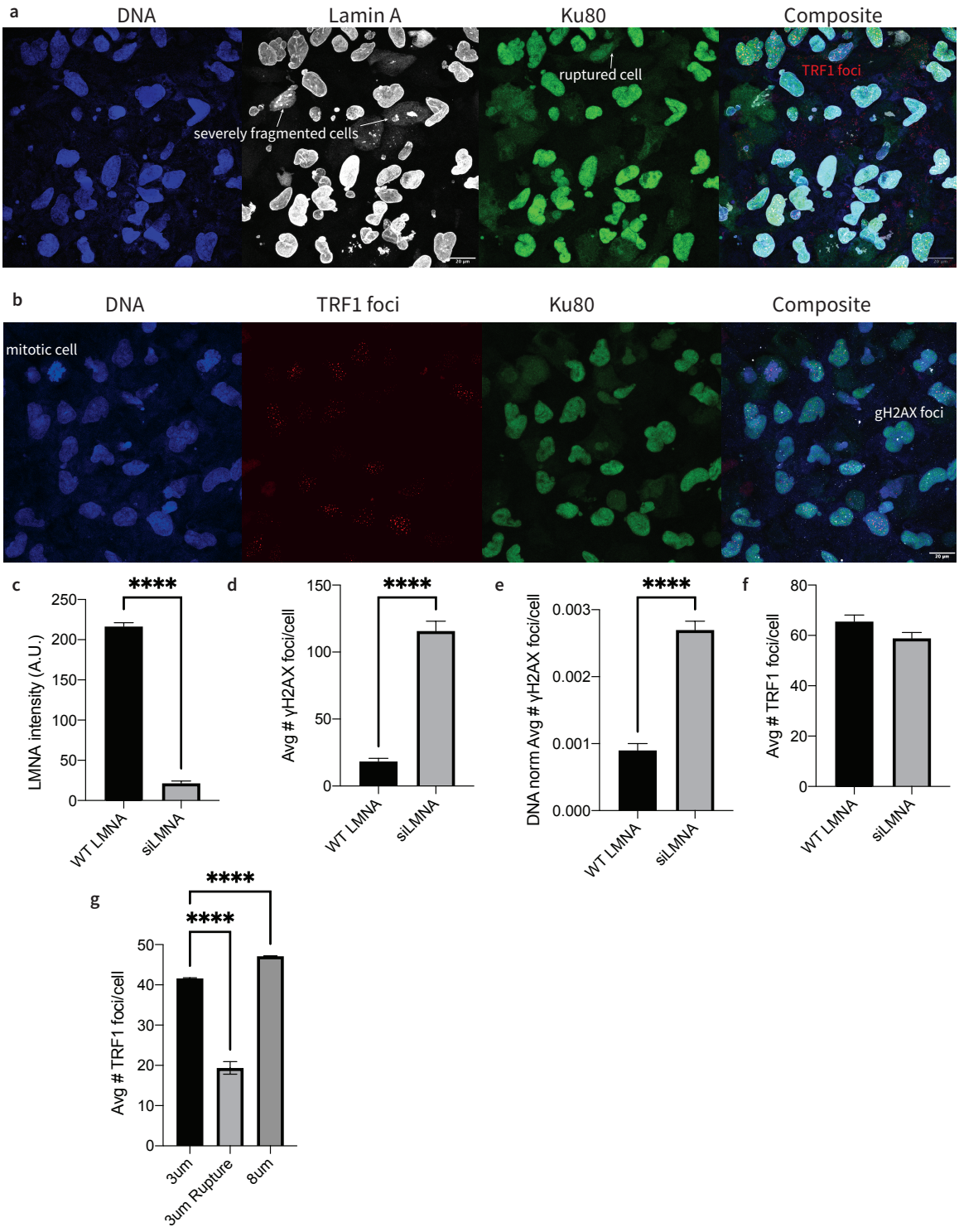
Figure 4-1 Acto-myosin stress inhibition and anti-oxidants reduce stress in a co-operative manner



4.2.2 Mis-localization of DNA repair factors during constricted migration leads to telomere attrition

We use ER inducible D450FOK1TRF1 U2OS cells where florescent TRF1 binds to the sheterin complex and indicated telomere ends as foci. We developed our own code to analyze the number and intensity if the TRF1 foci. U2OS cells has a nucleus ~20um which undergoes constricted migration through pores of 3um diameter resulting in some fragmented nuclei, nuclear membrane rupture and blebb formation (**Fig 4.2a**). Nonetheless, some cells continue to undergo mitosis after experiencing acute acto-myosin stress (**Fig 4.2b**). As studied in Chapter 2, Lamin A play a protective role against nuclear membrane rupture. U2OS cells with knockdown of Lamin A (siLMNA) (**Fig 4.2c**) show significantly greater gH2AX foci after undergoing constricted migration (**Fig 4.2d**). DNA damage is related to cell cycle state and increases during S phase of the cell cycle due to replication stress. To normalize for such effects, gH2AX is normalized to the total intensity of DNA of the cell, this shows significantly higher DNA damage in siLMNA U2OS as well (**Fig 4.2e**). The number of foci in siLMNA U2OS cells is lower (**Fig 4.2f**). Moreover, cells which are ruptured at the time of fixing, indicated by mis-localization of Ku80, show significantly lower number of TRF1 foci (**Fig 4.2g**) compared to other cells that do not currently have mis localized Ku80. Decrease in number of TRF1 foci also depends on the pore size of transwell used for constricted migration (Fig 4.2h), indicating TRF1 count is related probability of nuclear membrane rupture and curvature stress on the nucleus during constricted migration.

Figure 4-2 Lamin A protects against telomere attrition in cells undergoing constricted migration


a
DNA
Lamin A
Ku80
Composite
severely fragmented cells
ruptured cell
TRF1 foci
b
DNA
TRF1 foci
Ku80
Composite
mitotic cell
gH2AX foci
c
LMNA intensity (A.U.)
WT LMNA
siLMNA
****
d
Avg # γH2AX foci/cell
e
DNA norm Avg # γH2AX foci/cell
f
Avg # TRF1 foci/cell
g
Avg # TRF1 foci/cell
3um
3um Rupture
8um

Chapter 5 Big data analysis reveals single cell heterogeneity in cancer and embryonic tissue

SNPa, scRNA seq experiments and phenotype experiments in Fig 5.2f for A549 cells were done by Jerome Irianto and Yuntao Xia. scRNA seq and library prep for E5 embryonic chicken hearts was done by Mai Wang. Embryonic heart isolation was done by Manu Tewari.

5.1 Introduction

Aneuploidy/copy number variation (CNV) is a hallmark of cancer. Tumors can be thought of as ecosystems where each lineage of mutated malignant cell is 'selected' for in a Darwinian evolutionary process where one of the factors for 'natural' selection can also be therapy given to the patient. Mutations that provide therapy resistance to a particular set of malignant cells can lead to recurrence of cancer. Hence, it is important to track single cell variations in malignant cells. DNA/mRNA Sequencing technologies have rapidly evolved since the Human Genome Project was completed in 2003, state of the art single cell sequencing technology allows us to study heterogeneity in malignant cells. Gene dosage effects leading to phenotypic changes have been previously reported in yeast and mammalian cells(Kahlem *et al.*, 2004; Torres *et al.*, 2007; Williams *et al.*, 2008). Hence identifying underlying gene dosage effects can be vital to the study of emergence of malignancy as well as therapy resistance. There results show that besides epigenetic factors like transcription factors, histone modifications etc., aneuploidy can also play a major role in determining the phenotype of a cell. In this chapter we use constitutively expressed Lamin B1 to track spontaneously emerging CNV in live cancer cells and hypothesize that machine-learning dimensional reduction methods can aid discovering links between gene dosage and phenotype.

Tumor microenvironment can be heterogeneous among patients with the same type of cancer and can cause changes in tumor progression(Özdemir *et al.*, 2014; Patel *et al.*, 2014; Tirosh *et al.*, 2016). As illustrated in Chapter 3, single-cell sequencing allows us to not only parse out variation in malignant cells, but also identify other cells in the tumor microenvironment like cancer associated

fibroblasts (CAFs), tumor endothelial cells (TECs), tumor associated macrophages (TAMs) etc. In this chapter, we apply dimensional reduction techniques to identify constituent cell types in embryonic hearts and study the effect of drug perturbations in Chapter 2 at the single cell level.

Big data with dimensions of ~20000 genes and ~1000s cells is hard to visualize. Moreover, genes can show false 0 reads because of lack of sequencing depth due to prohibitive costs(Zhang, Ntranos and Tse, 2020). On average, only 30% of the genes in a single cell are sequenced. In this scenario, dimensional reduction techniques are important to identify data redundancy, clustering, and data visualization. Moreover, these methods prove useful in being able to identify unlabeled data, this is called unsupervised machine learning. Traditional linear methods of matrix factorization like principal component analysis (PCA), single value decomposition (SVD) etc allow as to significantly reduce the dimensions of the data matrix while capturing most of the variability in the data. However, state-of-the-art involves non-linear methods using neighbor graphs to recover the topology of the manifold and project it into lower dimensions, for example t-distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP). UMAP has advantages over t-SNE in being able to have 'meaningful' inter-cluster distance and reduced computational time(McInnes, Healy and Melville, 2018). In this chapter we focus on how these methods can be used to discover spontaneously arising heterogeneity within malignant cells, as well as diverse cell types in embryonic heart tissue.

5.2 Results

5.2.1 Machine Learning methods linking genotype to phenotype — using Constitutive protein Lamin B1 as a live cell marker to track Copy Number Variation

Previous chapters show that Lamin B1 is an 'essential' protein that is constitutively expressed in cells. We use aneuploid A549 cells in which 1 out of the 3 copies LMNB1 is tagged with florescent RFP to track loss of chromosome 5 in live cells (**Fig 5.1a**). Rare clonal colonies of cells in culture spontaneously lose florescence indicating loss of the copy of chromosome 5 that has RFP tagged Lamin B1 (**Fig 5.1b**). We select 4 representative RFP+ colonies and 3 RFP- colonies and expand the colonies to conduct single cell DNA sequencing. Single nucleotide polymorphism (SNP) array

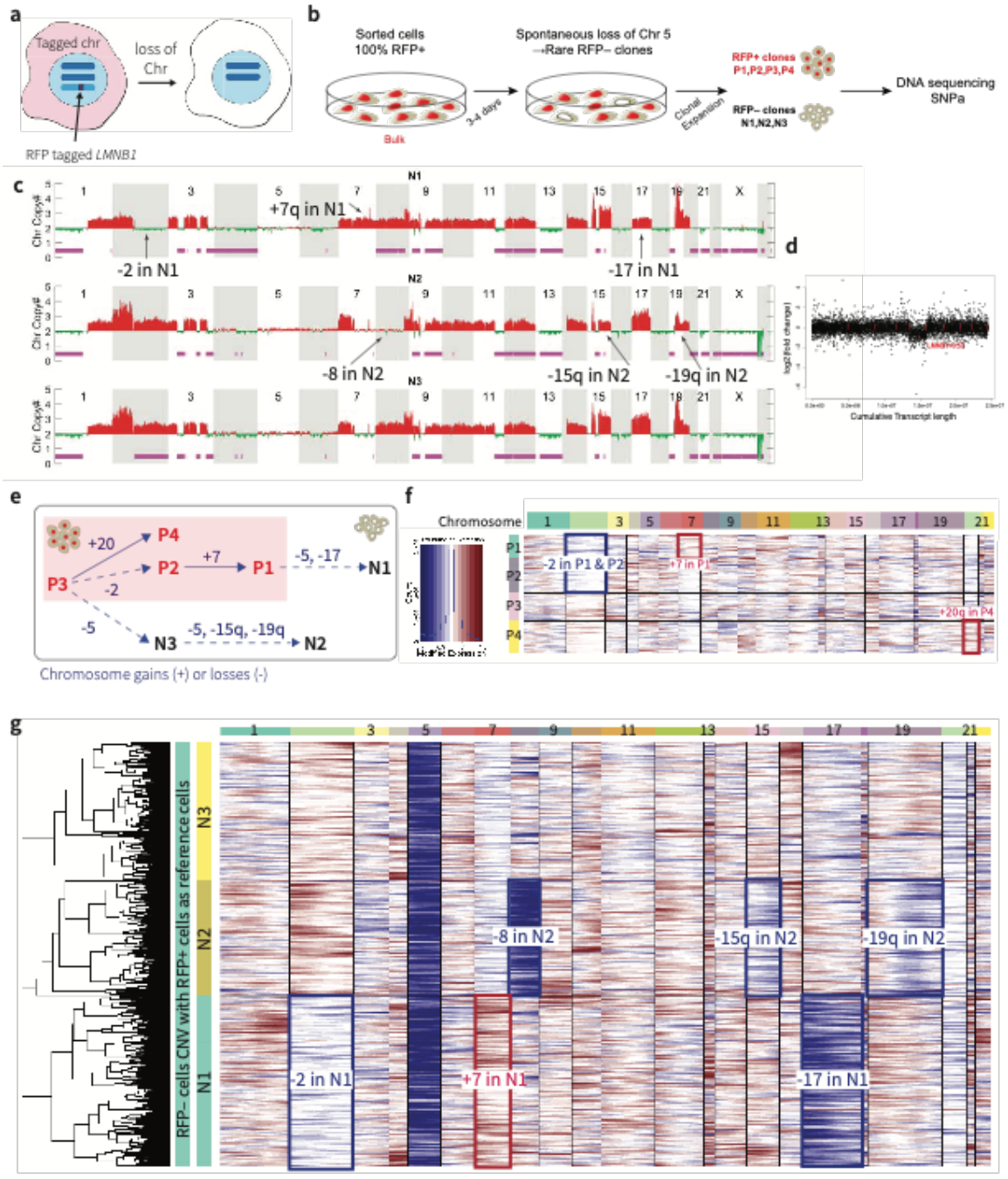
reveals that besides RFP- cells having only 2 copies of chromosome 5, they also have other CNV, for example N1 shows gain of chromosome 7 (**Fig 5.1c**).

5.2.1.1 InferCNV is used to identify CNV using scRNAseq

We now want to study if these CNV can cause gene dosage effects i.e can gain or loss of a chromosome on average affect the overall mRNA reads of the chromosome. We use 10x genomics single cell mRNA sequencing to get mRNA reads from all 4 RFP+ clones and all 3 RFP- clones. We find that dividing the average normalized reads of RFP+ cells from the RFP- shows reduced expression of chromosome 5 genes including *LMNB1* in RFP- cells (**Fig 5.1d**). Moreover, median factor change in expression of chromosome 5 genes in RFP- compared to RFP+ cells is 2/3 (**Fig S5-1a**), indicating gene dosage effect of losing 1 out of 3 copies of chromosome 5. Using SNPa data we can trace the lineage of the clones to find that P3 is the 'bulk' clone from which all the other cells have diverged by either gaining or losing an entire chromosome or p/q arm of a chromosome (**Fig 5.1e**). To study if each individual clone reveals gene dosage effects, we use inferCNV on unlabelled mRNA sequencing data of the clones. InferCNV(Patel *et al.*, 2014; Tirosh *et al.*, 2016) is an algorithm developed to parse out variability in mRNA reads coming from epigenetic factors, to those coming from gene dosage effects. To identify the unlabeled cells as (P1, P2, P3, P4 or N1, N2, N3) each dendrogram branch obtained from hierarchical clustering of InferCNV predicted chromosome gains/losses was matched to that from the SNPa data. Heatmap of InferCNV for RFP+ cells reveals the expected CNV from SNPa (**Fig 5.1f, Fig S5-1b**). Moreover, using RFP+ cells as reference cells, InferCNV predicts loss of Chromsome 5 across all 3 RFP- cells as well as predicting the CNV in each clonal population (**Fig 5.1g**).

Figure 5-1 Gene dosage effects linked to spontaneously arising CNV

a) Constitutive LMNB1 is fluorescently tagged to detect CNV in live cells b) Spontaneously arising RFP- and representative RFP+ cells are DNA sequenced c) SNPa confirms loss of chromosome 5 in RFP- cells and reveals other CNV d) Single cell mRNA sequencing reveals reduced reads from genes in chromosome 5 in RFP- cells e) Clone Lineage tracing derived through SNPa data f) Hierarchical clustering of InferCNV predicted CNV labels cells to the clones identified in SNPa. g) RFP- cells referenced to RFP+ cells reveal loss of chromosome 5 is predicted by mRNA reads as well.



5.2.1.2 UMAP differentiates clones identified by InferCNV and segregates based on gene dosage

Seurat(Butler *et al.*, 2018) was used to analyze 10x single-cell mRNA sequencing data performed on unlabeled RFP+ and RFP- cells. Quality control steps were performed to remove dead cells

(>12 % of reads are mitochondrial genes or number of unique genes mapped <500) and doublets/multiples of cells encompassed in the same droplet for 10x genomics (total reads > 6000/cell) (**Fig S5.1c,d**). scRNA data matrix is sparse as typically only 30% of the genes in a cell are sequenced. Hence it is important to reduce the dimensions of the matrix using PCA before conducting UMAP analysis on reduced matrix. Elbow plots are used to determine the number of PCs to consider for the UMAP analysis (**Fig S5.1e**). UMAP on dimensionally reduced matrix for single cell data is important. UMAP on the sparse data matrix with PCA dimensional reduction does not provide clusters that are as informative and resembles plots of PC1 vs other PCs (**Fig S5.1c,f**).

UMAP representation of RFP+ cells clusters all the cells of a particular clone together, confirming the clone labels obtained from hierarchical clustering using InferCNV (**Fig 5-2a**). Similarly for the RFP- clones (**Fig 5.2b**). Identifying differentially expressed genes that show >1.28 times the expression in RFP+ cells compared to RFP- cells, most of those genes belong to chromosome 5 (**Fig 5.2c**). Chromosome 5 has >7 times the average number of RFP+ upregulated genes/chromosome (average=3). This shows a gene dosage effect in the RFP+ cells.

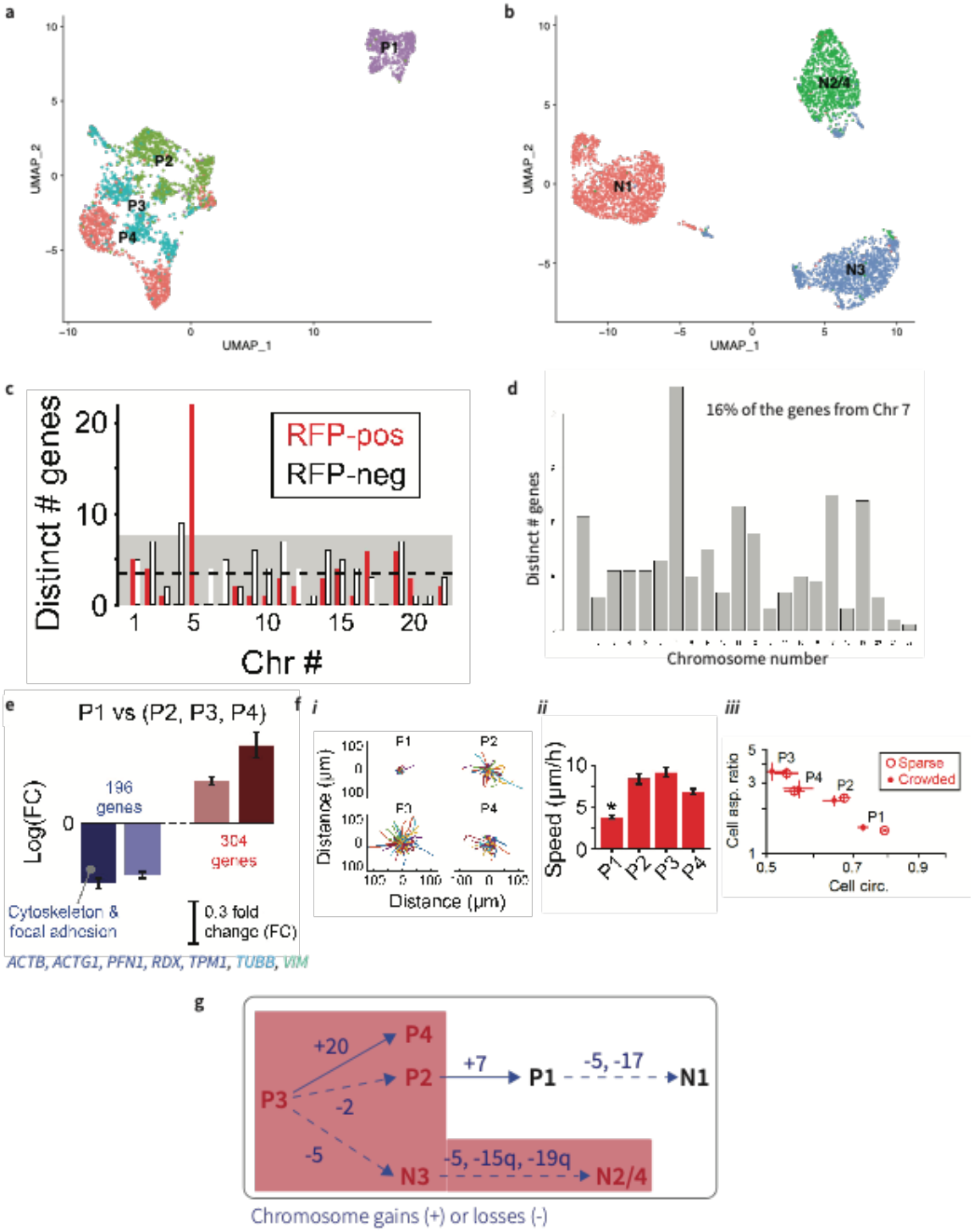
Interestingly, UMAP of RFP+ cells shows that P1 which has gain of chromosome 7, clusters away from the other 3 lineages. This is also reflected a gene dosage response where a disproportionate number of differentially upregulated genes in P1 compared to (P2, P3, P4) are from chromosome 7 (**Fig 5.2d**). Gene Ontology (GO) Annotation of the 196 genes that are differentially downregulated in P1 enrich for the GO terms cytoskeleton and focal adhesion, including genes like *ACTB, ACTG1, PFN1, RDX, TPM1, TUBB, VIM* (**Fig 5.2e**).

Phenotypic changes are observed in P1 lineage compared to (P2, P3, P4). P1 cells show decrease cell migration persistence, decreased cell speed and increased cell circularity (**Fig 5.2f**). These phenotypic changes can be attributed to differential mRNA expression given that cytoskeleton plays a key role in cell motility and focal adhesions play a key role in cell spreading (**Fig 5.2g**).

Figure 5-2 UMAP differentiates clones identified by InferCNV and segregates based on gene dosage

UMAP clusters together clones identified by InferCNV in **a)** RFP+ and **b)** RFP- cells. **c)** Differential expression of genes in RFP+ compared to RFP- cells reveals 7 times the average number of RFP+

upregulated genes are in chromosome 5. **d)** Similar analysis comparing P1 to (P2, P3, P4) reveals most P1 upregulated genes are in chromosome 7. **e)** GO term enrichment reveals most downregulated genes annotate for cytoskeleton and focal adhesions. **f)** Phenotypic differences observed in P1 cells include reduced ***i)*** cell persistence, ***ii)*** cell speed and ***iii)*** cell spreading. **g)** Phenotypic differences can arise from spontaneous CNV in genetically unstable cancer cells



5.2.1.3 Unsupervised machine learning techniques capture and predict the differential phenotype observed

Gene	Avg logFC	pct.1	pct.2
VIM	-1.34	0.982	0.999
TGFBI	-1.24	0.926	0.987
LINC00152	-1.23	0.6	0.966
IGFBP7	-1.11	0.638	0.953
ACTG1	-1.11	0.998	1
TPM1	-1.08	0.971	0.993
EIF1	-1.02	1	0.999
PMEPA1	-0.96	0.843	0.933
NPC2	-0.88	0.984	0.997
KRT7	-0.85	0.964	0.974
RSPO3	-0.85	0.345	0.899
MIR4435-2HG	-0.80	0.403	0.866
NPM1	-0.79	0.998	0.998
IGFBP4	-0.78	0.991	0.997
RPS7	-0.75	1	1

Table 5-1 *15 genes that show the most downregulation in P1 cells compared to other lineages (pct.1=% cells in P1 expressing the gene; pct.2=% cells in P2, P3, P4 expressing the gene)*

Top 15 downregulated genes in P1 compared to (P2, P3, P4) include vimentin and long non-coding RNA *LINC00152* and *MIR4435-2HG* (**Fig 5.3a**). Vimentin is involved in cell migration, motility, adhesion and is a widely used marker for epithelial to mesenchymal transition (EMT)(Mendez, Kojima and Goldman, 2010). EMT is an important process for cancer metastasis in which adhesive epithelial cells become migratory. Downregulation of Vimentin by P1 cells might indicate mesenchymal to epithelial transition (MET). Moreover, this is also reflected in the phenotypic changes observed in **Fig 5.2f**). P1 upregulated genes annotate for Type 1 interferon response (**Fig 5.3b**). Principal Component Analysis (PCA) is a matrix factorization technique that helps reduce

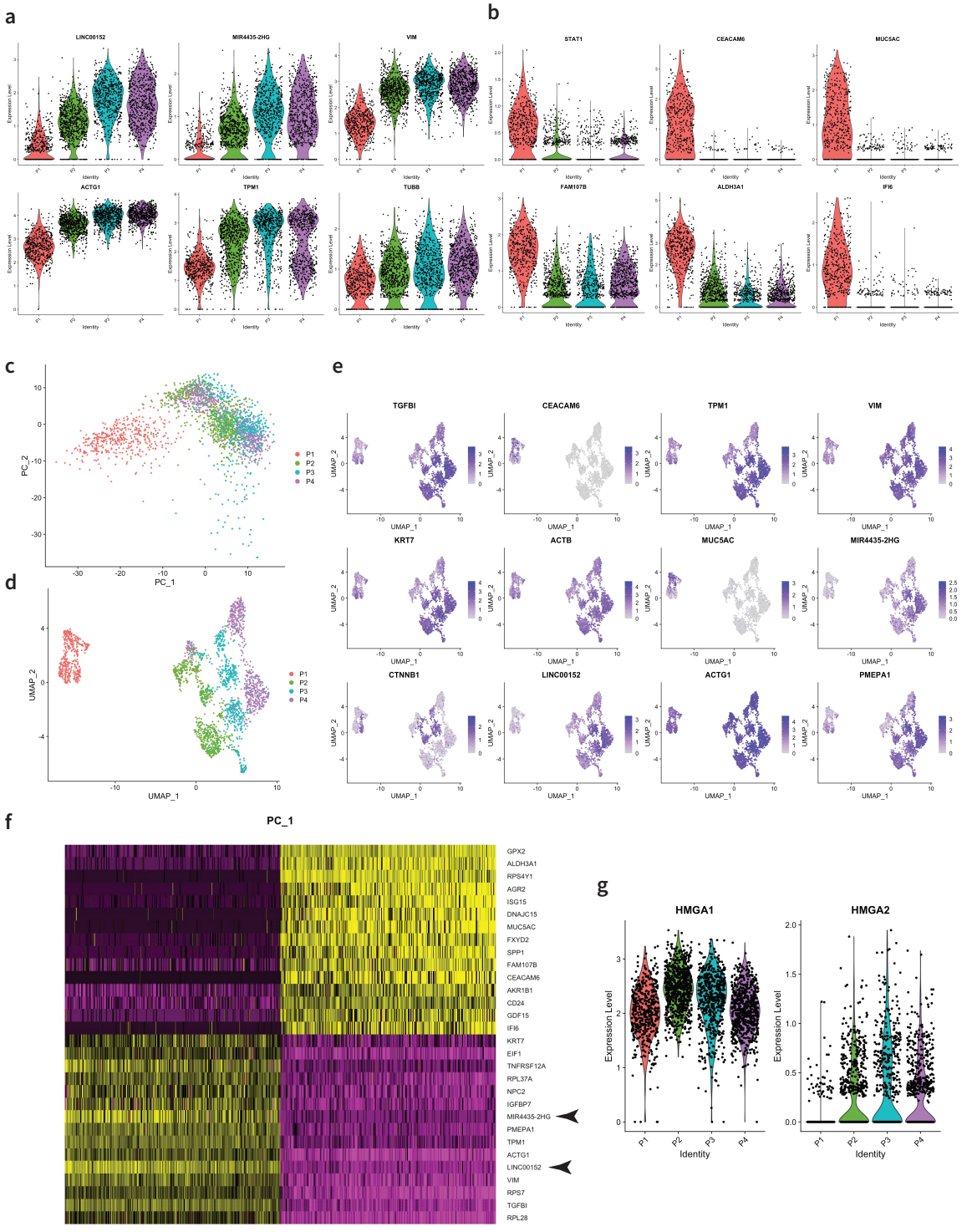
the dimensions of data by considering only eigenvectors of the covariance matrix that have eigenvalues are above a set threshold. **Fig S5-1e** shows the first principal component (PC) captures most of the variance in the data matrix of genes vs RFP+ cells. Plotting the PC1 vs PC2 shows that P1 cells separate from the other cells based on PC1 (**Fig 5.3c**). Hence PCA predicts difference in gene expression of P1 cells compared to others. This is also observed in UMAP representation of RFP+ cells showing P1 cells clustering away from the rest and showing low reads of cell motility genes (**Fig 5.3d,e**). Downregulation of *VIM*, *LINC00152* and *MIR4435-2HG* also define PC1 which is rationalized by the downregulation analysis above (**Fig 5.3f**).

Long non-coding RNA are a sequence of >200 bp that do not map to any protein sequence. However, they can have many functions. Sequence specific structure of *LINC00152* has been shown to promote invasive migratory phenotype in glioblastoma(Reon *et al.*, 2018). *LINC00152* has been shown to promote EMT in glioblastoma through enhanced expression of HMGA2 (X *et al.*, 2018), which is also found to be downregulated in P1 (**Fig 5.4g**). *MIR4435-2HG* has been shown to maintain EMT in lung cancer cell lines through preventing proteosomal degradation of β-catenin(Qian *et al.*, 2018).

Hence unsupervised machine learning techniques capture and predict the differential phenotype observed in P1 cells.

Figure 5-3 P1 downregulates EMT and loses cell motility

Gene markers a) upregulated and b) downregulated in P1 cells. c) PC1 separates P1 genetic profile from the rest. d) Differential gene expression of the P1 gene markers is also observed in the spatial representation of the UMAP f) Top 500 cells and 15 genes that define PC1 are represented as a heatmap where yellow represents high reads of the gene in that cell g) HMGA1 and HMGA2 are downregulated in P1 cells

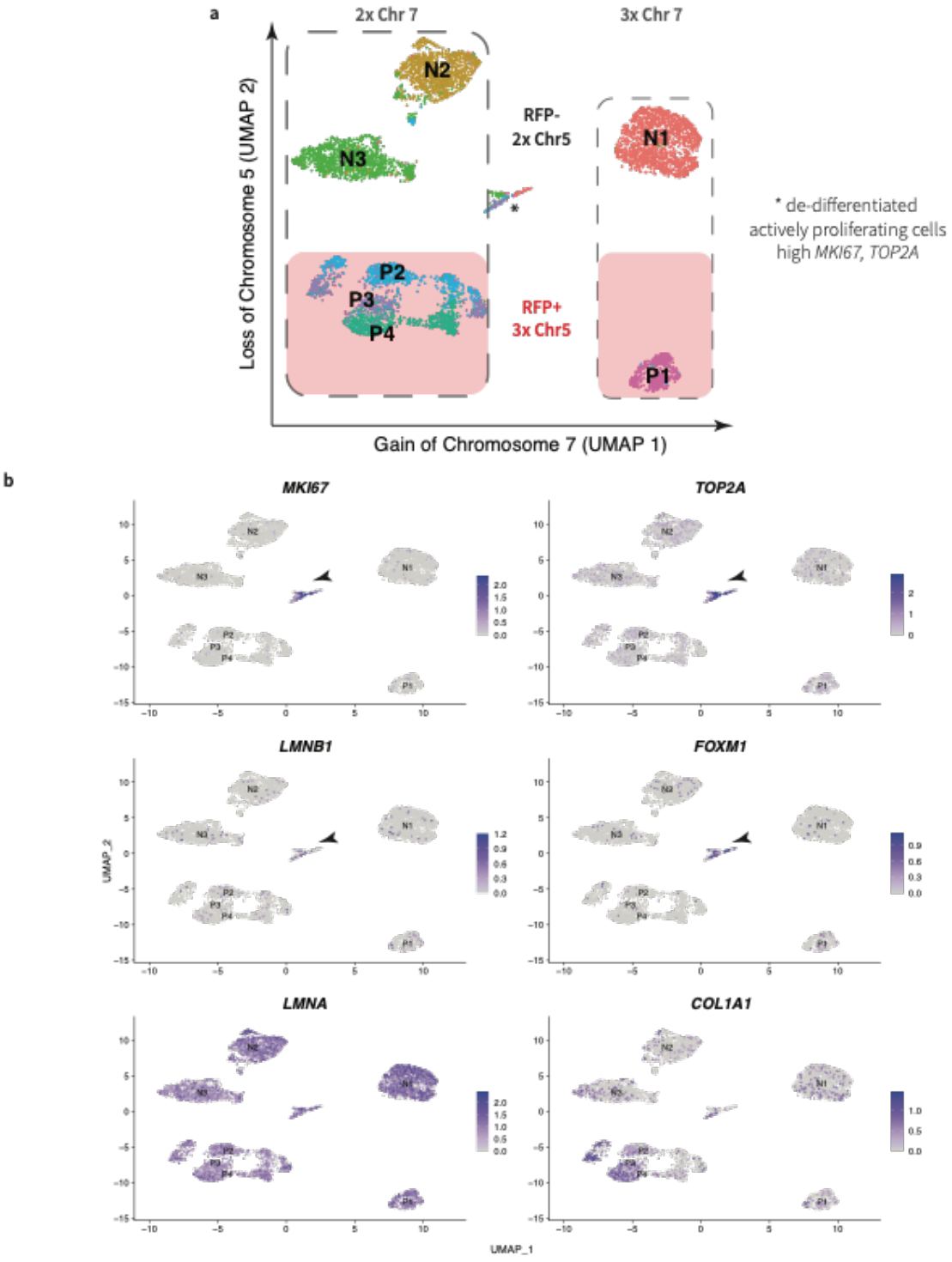


UMAP analysis on merged data matrix of RFP+ and RFP- cells shows UMAP 1 captures differences in mRNA expression due to gain in chromosome 7, whereas UMAP 2 captures differences in mRNA expression due to loss of chromosome 5 (**Fig 5.4a**). Moreover, in the center

is a small cluster of actively proliferating cells indicated by their high reads of *MKI67, TOP2A, LMNB1, FOXM1* (**Fig 5.4b**). This might indicate that actively proliferating cancer cells are 'de-differentiated' and cluster together irrespective of their lineage.

Figure 5-4 UMAP captures and predicts differential phenotype observed

a) Gain of chromosome 7 and loss of chromosome 5 is represented in the UMAP axes **b)** Heatmap representation of cells in UMAP show 'de-differentiated' cluster of actively proliferating cells based on high reads of *MKI67, TOP2A, LMNB1, FOXM1*



5.2.2 Machine Learning methods identify constitutive cell types in embryonic heart tissue

Chapter 2 studies the role of Lamin-A as a mechanosensitive protector against DNA damage and cell cycle arrest in embryonic hearts. However, the embryonic heart tissue is a conglomeration of

different cell types, and we would like to further study if the acto-myosin stress inhibition and DNA damage can have cell type dependent effects. To do that, we must first identify the cell types. Here we aim to use dimensional reduction methods developed in the previous section to identify constitutive cell types in embryonic heart tissue.

We performed 10x single cell mRNA sequencing on 4 types of samples from Chapter 2- DMSO (Ctrl), Blebb (2 hrs of blebbistatin treatment), BlebbWO (2 hrs of blebb treatment followed by 1 hr of washout), H_2O_2 (3 hrs of oxidative DNA damage). We first do quality control checks on the sample and remove any cells that have >12% mitochondrial genes and <200 total reads (indicating that they might be dead/empty droplets). We also remove cells with higher that >3000 reads which might be doublets/multiplets of cells (**Fig S5.2a**). Principal components analysis (PCA) enables us to reduce the dimensions of the sparse matrix of genes vs cells. PCs identify the orthogonal vectors along which the cells show the most variation, **Fig S5.2b** lists the genes that define the PCs. We use the Elbow plot to determine that reducing the dimensions of the matrix to 15 PCs captures most of the variation in the matrix (**Fig S5.2c**).

UMAP of the reduced matrix shows that samples do not cluster together, instead cell types common across the samples cluster together (**Fig 5.5a**). The only exception to this is the DMSO sample has cells that exclusively dominate cluster number 5. We use differential expression analysis to define cluster markers that show >1.28 times the expression in that cluster compared to others. Three clusters show high reads for fibrous ECM genes - *COL1A1, COL1A2, COL3A1* indicating epithelial or epicardium phenotype (**Fig S5.3a**). Development of heart valve cells involves mesenchymal cells that express *SOX9* which further upregulates chondrogenesis(Garside *et al.*, 2015). Hence, we label this cluster as valve cells (**Fig S5.3b**). The cluster that contains only DMSO cells shows high reads of ECM genes as well as *CRABP1* which is downstream of retinoic acid signaling that regulates *LMNA* expression (Swift *et al.*, 2013). We label this cluster as DMSO epicardial cells, which perhaps are only present in hearts that have not undergone acto-myosin inhibition or extensive DNA damage (**Fig S5.3c**). Cardiomyocytes were identified by high reads of troponin and actinin reads (**Fig S5.4a**). Endothelial cells were identified by high reads of *PECAM1*, cadherin and

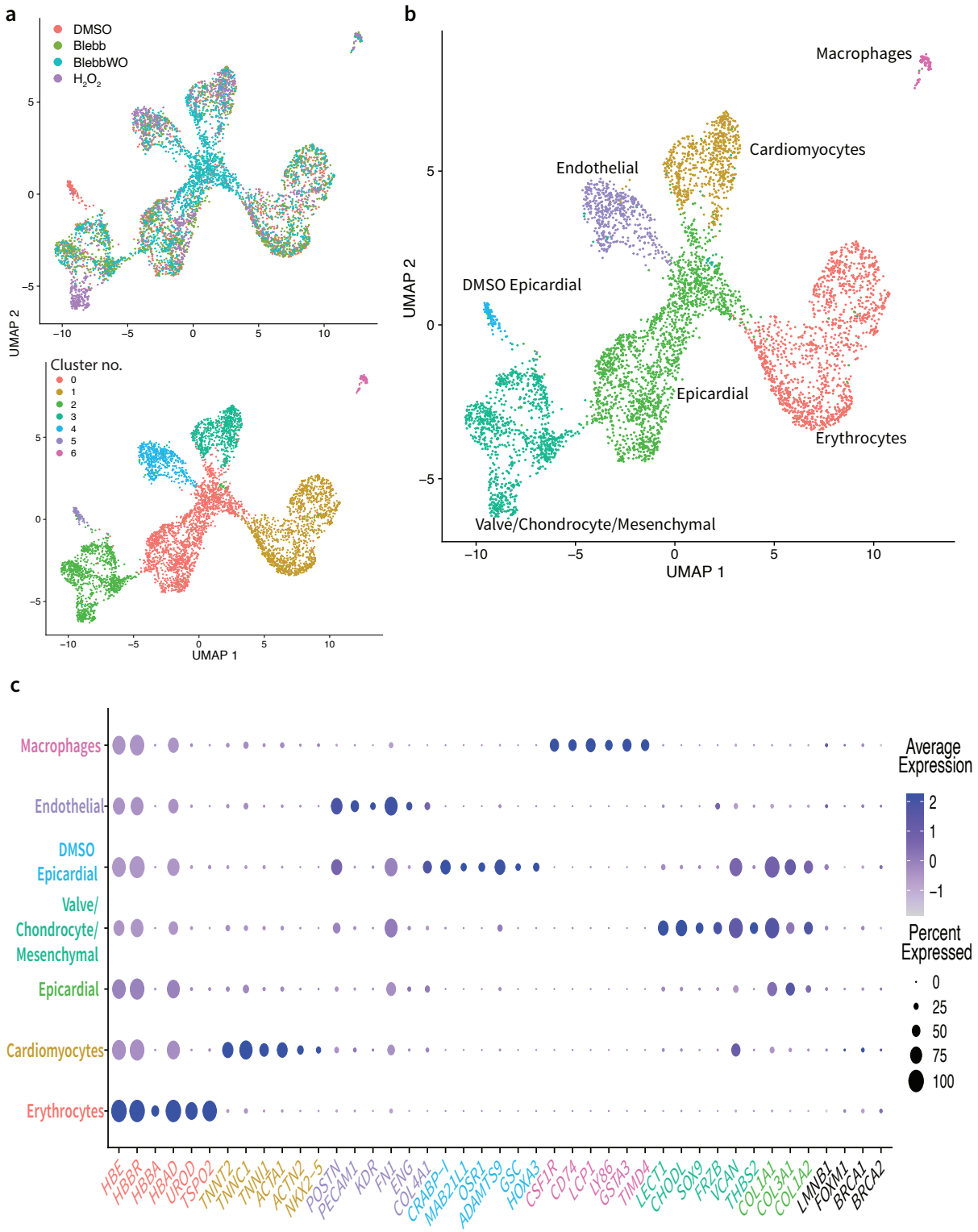
angiogenesis genes (**Fig S5.4b**). Erythrocytes were identified by high reads of heme-globin and heme biosynthesis genes (**Fig S5.4c**). Lastly macrophages were identified by markers like *CD74, CSF1R* (**Fig S5.4d**).

UMAP successfully segregates cell types into distinct clusters (**Fig 5.5b**), which can be identified by high reads of specific cell-type markers (**Fig 5.5c**).

Each drug perturbation has roughly same proportion of cell types, except for the control (DMSO) sample also having specific epicardial cell population with distinct markers (**Fig 5.6a**). Among the cells sequenced, cells clustered as epicardial cells show the highest number (**Fig 5.6b**). In contrast to other cell types, erythrocytes show a higher number of reads for the number of unique genes that the reads are mapped to, perhaps indicating advanced state of differentiation - high reads of a limited number of genes (**Fig 5.6c**). This has also been observed in published scRNA data from chicken embryonic hearts(M *et al.*, 2021). The alternative hypothesis is that they are droplets that also had lysed erythrocytes or floating mRNA along with the cell being sequenced, resulting in high reads of heme biogenesis genes.

Figure 5-5 UMAP clusters cell type across hearts with different drug perturbations

a) UMAP does not segregate cells based on drug perturbation **b)** but by constituent cell types **c)** which can be identified by gene expression profile specific to the differentiated cells


a
DMSO
Blebb
BlebbWO
H2O2
UMAP 2
Cluster no.
0
1
2
3
4
5
6
UMAP 1
b
Macrophages
Cardiomyocytes
Endothelial
DMSO Epicardial
Epicardial
Erythrocytes
Valve/Chondrocyte/Mesenchymal
UMAP 2
UMAP 1
c
Macrophages
Endothelial
DMSO Epicardial
Valve/ Chondrocyte/ Mesenchymal
Epicardial
Cardiomyocytes
Erythrocytes
HBE
HBBR
HBBA
HBAD
UROD
TSPO2
TNNT2
TNNC1
TNNI1
ACTA1
ACTN2
NKX2-5
POSTN
PECAM1
KDR
FN1
ENG
COL4A1
CRABP-I
MAB21L1
OSR1
ADAMTS9
GSC
HOXA3
CSF1R
CD74
LCP1
LY86
GSTA3
TIMD4
LECT1
CHODL
SOX9
FRZB
VCAN
THBS2
COL1A1
COL3A1
COL1A2
LMNB1
FOXM1
BRCA1
BRCA2
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
100

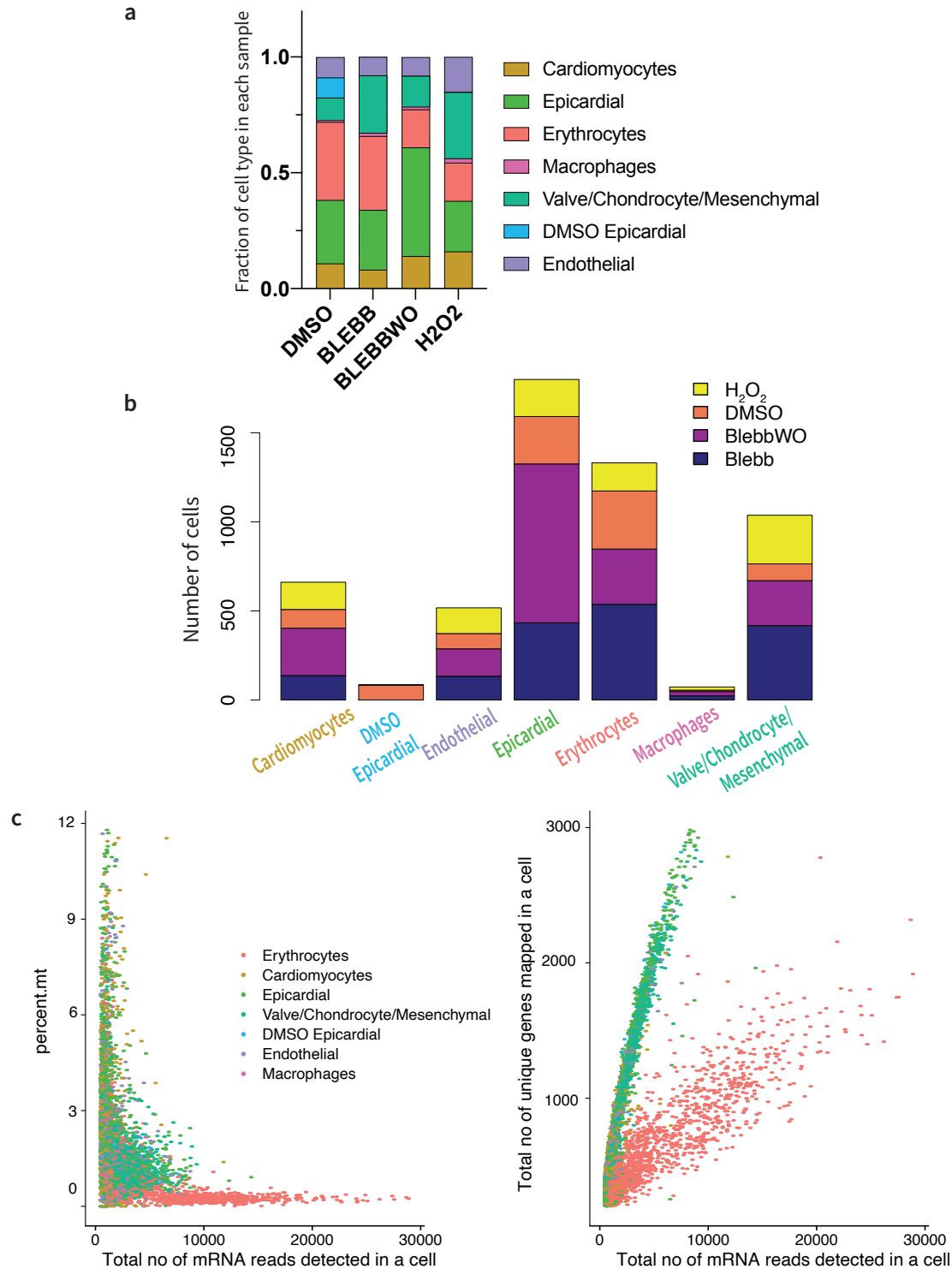


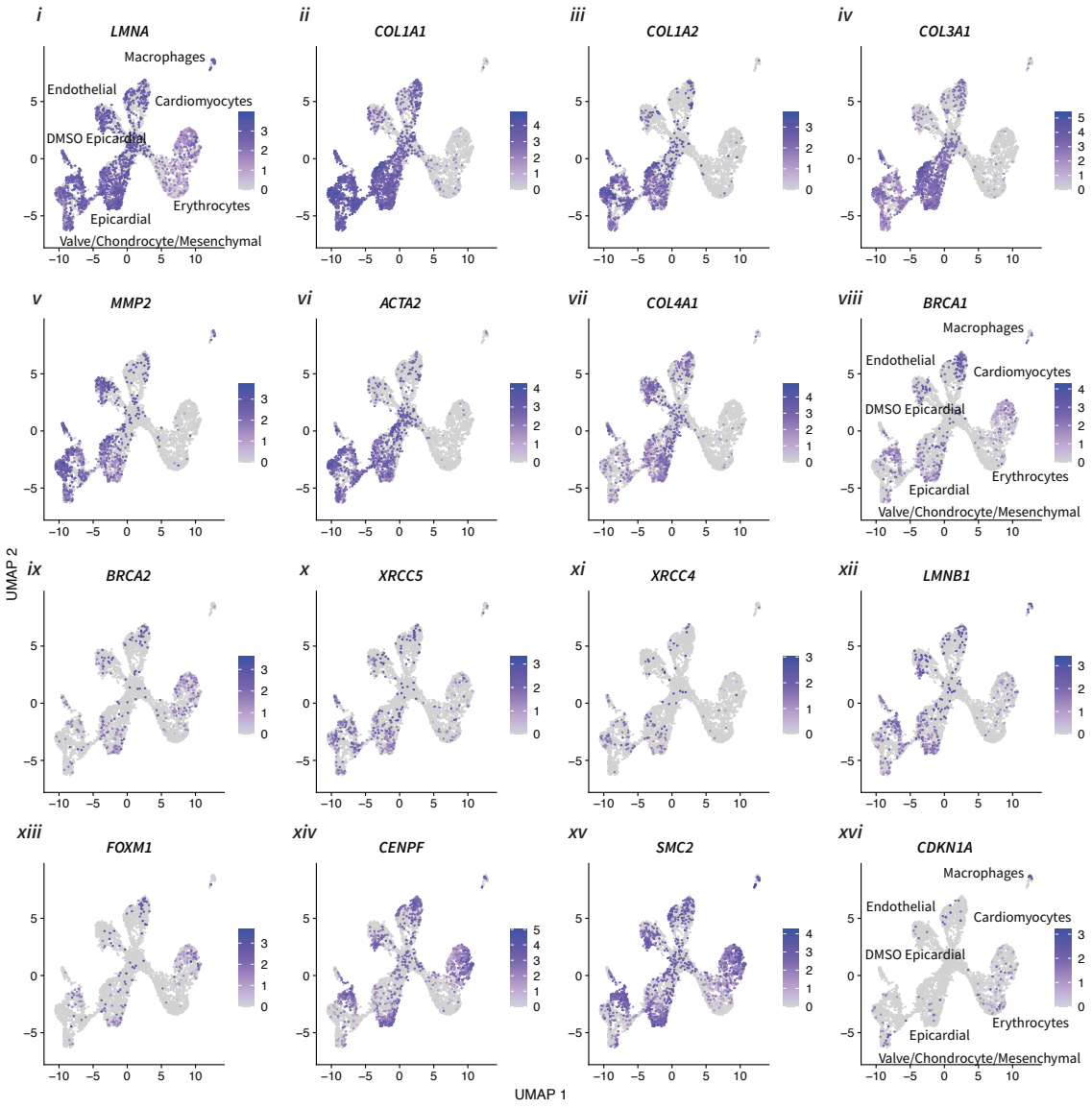
Figure 5-6 Distribution of cell types in hearts treated with different drug perturbations

a) Cell type count normalized to total number of cells in the sample **b)** Distribution of cell type across samples **c)** Total count of mRNA reads compared to total number of unique genes mapped in a cell shows erythrocytes have a disproportionately high total count of mRNA

5.3 Discussion

LMNA shows high reads across cell types (**Fig 5.7*i***). Genes coding for heterodimers of collagen-1, *COL1A1* and *COL1A2* show high reads in concurrent cells, as does another fibrous ECM gene - *COL3A1* (**Fig 5.7*ii-iv***). Two major double strand break DNA repair pathways are Homologous Recombination (HR) and Non-homologous end joining (NHEJ). HR genes (*BRCA1, BRCA2*) and NHEJ genes (*XRCC5, XRCC4*) show very few high reads that equitably distributed across cell types (**Fig 5.7*vii-ix***). *LMNB1* shows high reads in cells that are actively proliferating as indicated by the concurrent high reads of *FOXM1, CENPF, SMC2*. In contrast, *LMNA* shows high reads uniformly across diverse cells. It is also interesting to note that unlike cancer cells in culture, actively proliferating cells in embryonic heart tissue are distributed throughout cell types and do not 'de-differentiate' to form one cluster (**Fig 5.7*xii-xv***). Very few high reads of p21 (*CDKN1A*) are observed, which is downstream of p53 cell cycle arrest pathway. Interestingly epicardial cells specific to the DMSO (Ctrl) sample do not show any high reads of *CDKN1A* (**Fig 5.7*xvi***).

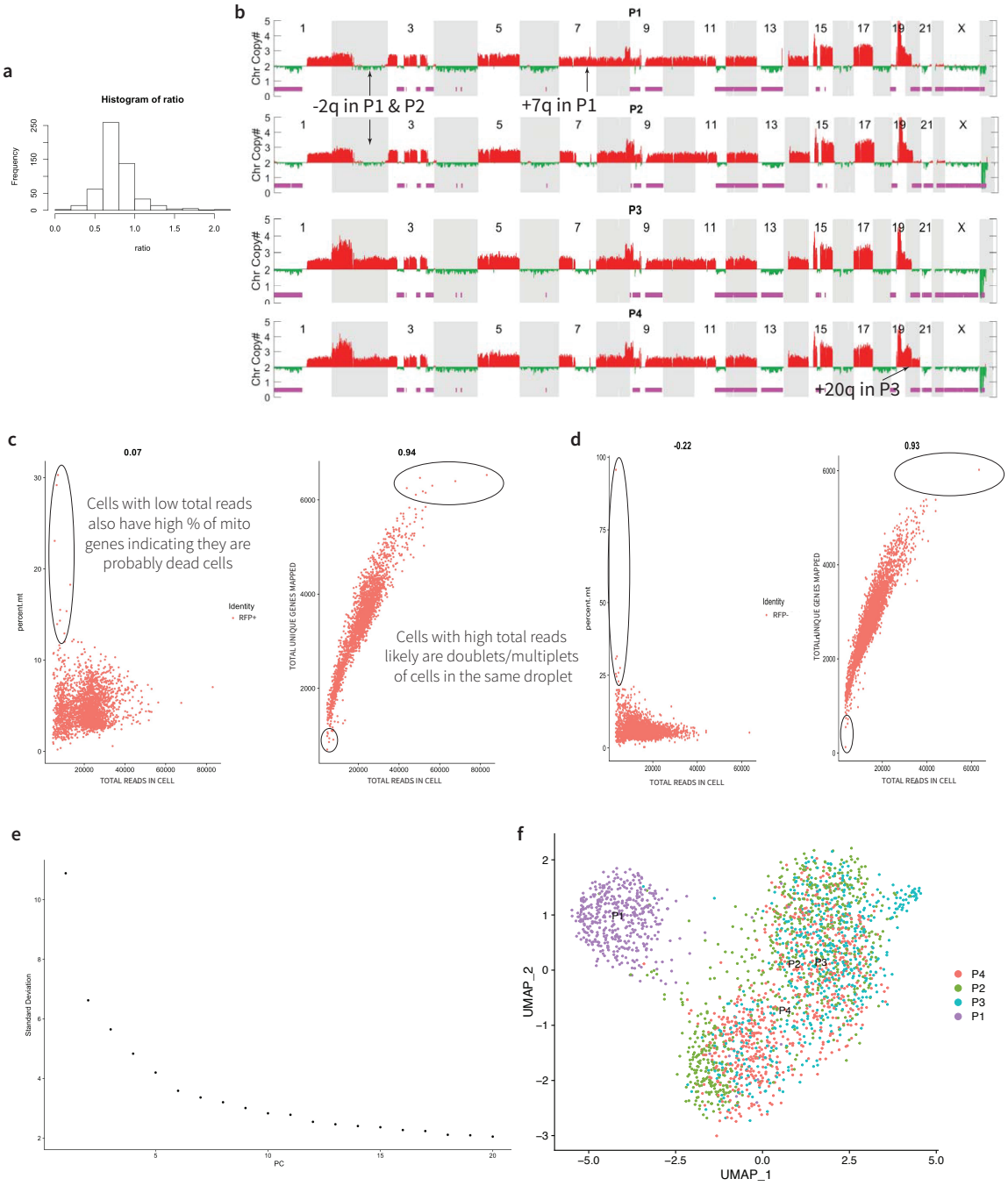
Figure 5-7 Heatmap based on the reads of genes in individual cells spatially organized in UMAP clusters

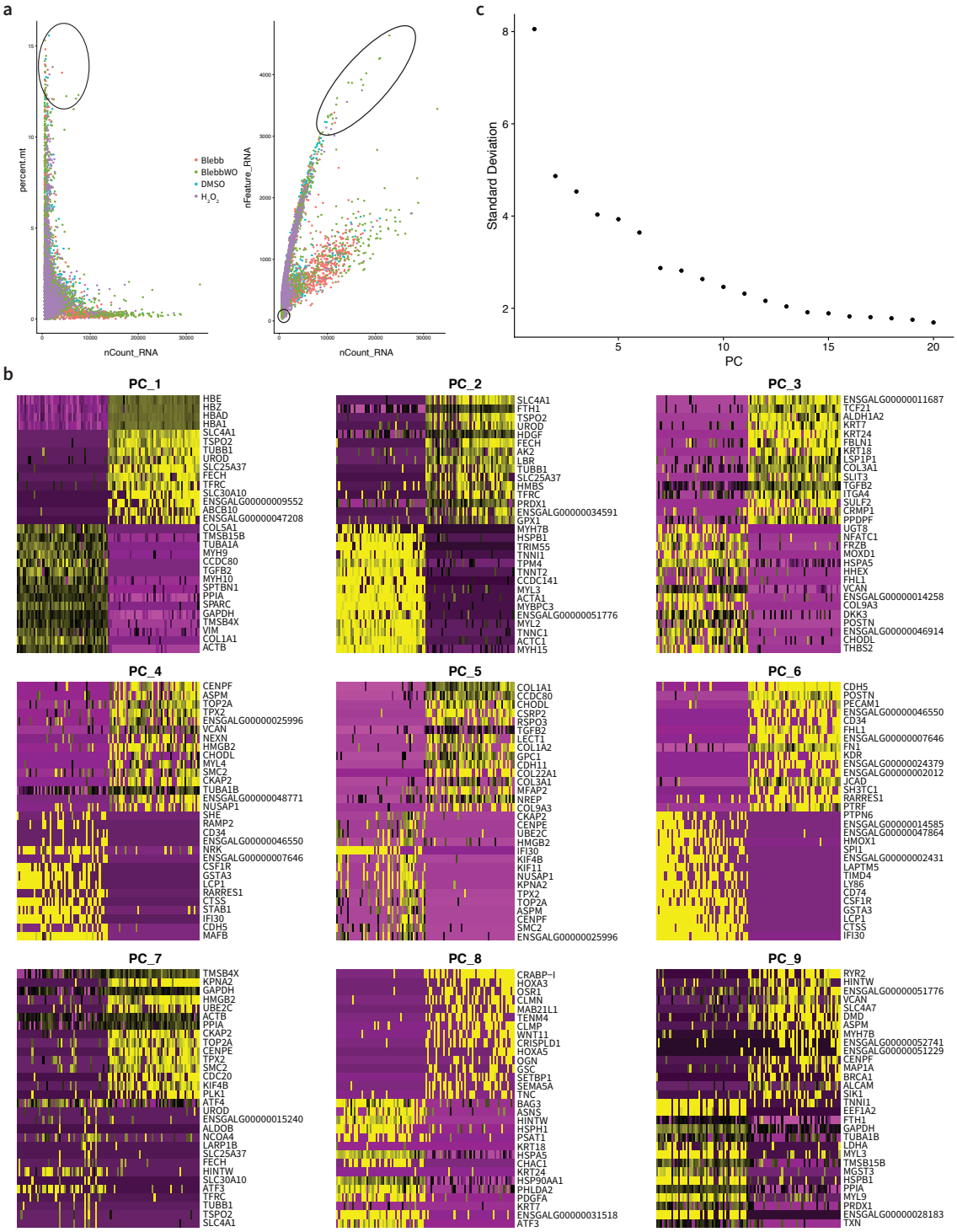


3.4 Supplementary Figures

Fig S5.1 a) 0.67 is the most common fold change in chromosome 5 gene expression of RFP-/RFP+ A549 cells **b)** SNPa of 4 divergent clones of RFP+ A549 cells. Cells with high % of mitochondrial genes or low total mRNA reads were removed for **c)** RFP+ and **d)** RFP- cells. **e)** Elbow plots were used to determine the number of PCs to include for further analysis. **f)** UMAP on data matrix without dimensional reduction resembles PC1 vs PC2 plots

Fig S5.2 a) Cells with high % of mitochondrial genes or low total mRNA reads were removed as they are probably dead cells/empty droplets **b)** Top 15 upregulated and downregulated genes that define principal component vectors and heatmap of the top 100 cells that show high reads (in yellow) of upregulated genes and low reads (pink) of downregulated genes (pink) **c)** Elbow plots were used to determine the number of PCs to include for further analysis


a
Histogram of ratio
Frequency
ratio
b
P1
P2
P3
P4
Chr Copy#
-2q in P1 & P2
+7q in P1
+20q in P3
c
0.07
Cells with low total reads also have high % of mito genes indicating they are probably dead cells
percent.mt
Identity
RFP+
TOTAL READS IN CELL
0.94
Cells with high total reads likely are doublets/multiplets of cells in the same droplet
TOTAL UNIQUE GENES MAPPED
d
-0.22
RFP-
0.93
e
Standard Deviation
PC
f
UMAP_2
UMAP_1
P4
P2
P3
P1


a
percent.mt
nCount_RNA
Blebb
BlebbWO
DMSO
H2O2
nFeature_RNA
nCount_RNA
c
Standard Deviation
PC
b
PC_1
HBE
HBZ
HBAD
HBA1
SLC4A1
TSPO2
TUBB1
UROD
SLC25A37
FECH
TFRC
SLC30A10
ENSGALG00000009552
ABCB10
ENSGALG00000047208
COL5A1
TMSB15B
TUBA1A
MYH9
CCDC80
TGFB2
MYH10
SPTBN1
PPIA
SPARC
GAPDH
TMSB4X
VIM
COL1A1
ACTB
PC_2
SLC4A1
FTH1
TSPO2
UROD
HDGF
FECH
AK2
LBR
TUBB1
SLC25A37
HMBS
TFRC
PRDX1
ENSGALG00000034591
GPX1
MYH7B
HSPB1
TRIM55
TNNI1
TPM4
TNNT2
CCDC141
MYL3
ACTA1
MYBPC3
ENSGALG00000051776
MYL2
TNNC1
ACTC1
MYH15
PC_3
ENSGALG00000011687
TCF21
ALDH1A2
KRT7
KRT24
FBLN1
KRT18
LSP1P1
COL3A1
SLIT3
TGFB2
ITGA4
SULF2
CRMP1
PPDPF
UGT8
NFATC1
FRZB
MOXD1
HSPA5
HHEX
FHL1
VCAN
ENSGALG00000014258
COL9A3
DKK3
POSTN
ENSGALG00000046914
CHODL
THBS2
PC_4
CENPF
ASPM
TOP2A
TPX2
ENSGALG00000025996
VCAN
NEXN
HMGB2
CHODL
MYL4
SMC2
CKAP2
TUBA1B
ENSGALG00000048771
NUSAP1
SHE
RAMP2
CD34
ENSGALG00000046550
NRK
ENSGALG00000007646
CSF1R
GSTA3
LCP1
RARRES1
CTSS
STAB1
IFI30
CDH5
MAFB
PC_5
COL1A1
CDC80
CHODL
CSRP2
RSPO3
TGFB2
LECT1
COL1A2
GPC1
CDH11
COL22A1
COL3A1
MFAP2
NREP
COL9A3
CKAP2
CENPE
UBE2C
HMGB2
IFI30
KIF4B
KIF11
NUSAP1
KPNA2
TPX2
TOP2A
ASPM
CENPF
SMC2
ENSGALG00000025996
PC_6
CDH5
POSTN
PECAM1
ENSGALG00000046550
CD34
FHL1
ENSGALG00000007646
FN1
KDR
ENSGALG00000024379
ENSGALG00000002012
JCAD
SH3TC1
RARRES1
PTRF
PTPN6
ENSGALG00000014585
ENSGALG00000047864
HMOX1
SPI1
ENSGALG00000002431
LAPTM5
TIMD4
LY86
CD74
CSF1R
GSTA3
LCP1
CTSS
IFI30
PC_7
TMSB4X
KPNA2
GAPDH
HMGB2
UBE2C
ACTB
PPIA
CKAP2
TOP2A
CENPE
TPX2
SMC2
CDC20
KIF4B
PLK1
ATF4
UROD
ENSGALG00000015240
ALDOB
NCOA4
LARP1B
SLC25A37
FECH
HINTW
SLC30A10
ATF3
TFRC
TUBB1
TSPO2
SLC4A1
PC_8
CRABP-I
HOXA3
OSR1
CLMN
MAB21L1
TENM4
CLMP
WNT11
CRISPLD1
HOXA5
OGN
GSC
SETBP1
SEMA5A
TNC
BAG3
ASNS
HINTW
HSPH1
PSAT1
KRT18
HSPA5
CHAC1
KRT24
HSP90AA1
PHLDA2
PDGFA
KRT7
ENSGALG00000031518
ATF3
PC_9
RYR2
HINTW
ENSGALG00000051776
VCAN
SLC4A7
DMD
ASPM
MYH7B
ENSGALG00000052741
ENSGALG00000051229
CENPF
MAP1A
BRCA1
ALCAM
SIK1
TNNI1
EEF1A2
FTH1
GAPDH
TUBA1B
LDHA
MYL3
TMSB15B
MGST3
HSPB1
PPIA
MYL9
PRDX1
ENSGALG00000028183
TXN

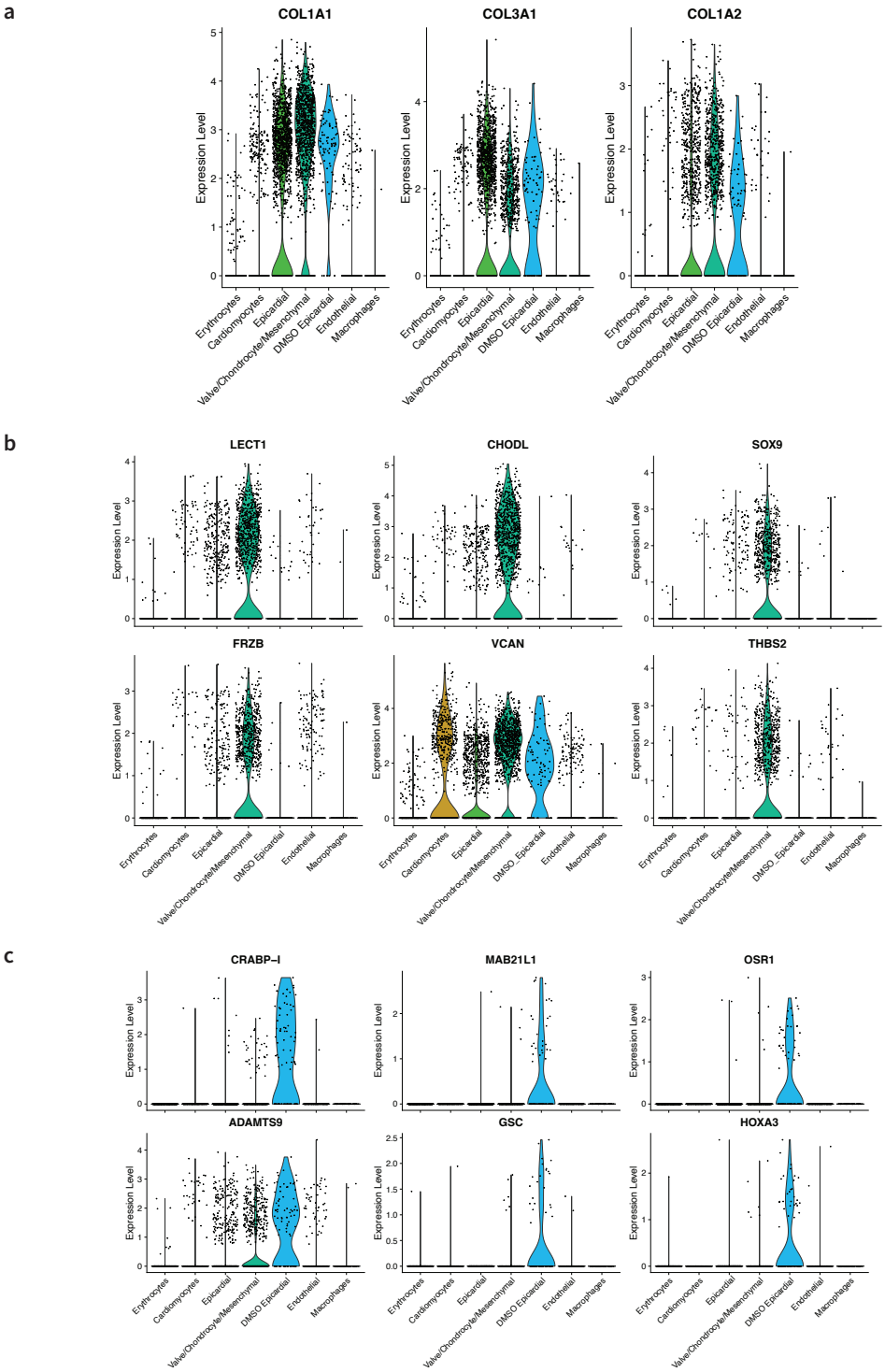


Fig S5.3 Cell type specific expression of **a)** fibrous ECM genes **b)** Chondrogenesis genes and *SOX9* **c)** Genes upregulated in DMSO (Ctrl) sample specific cells

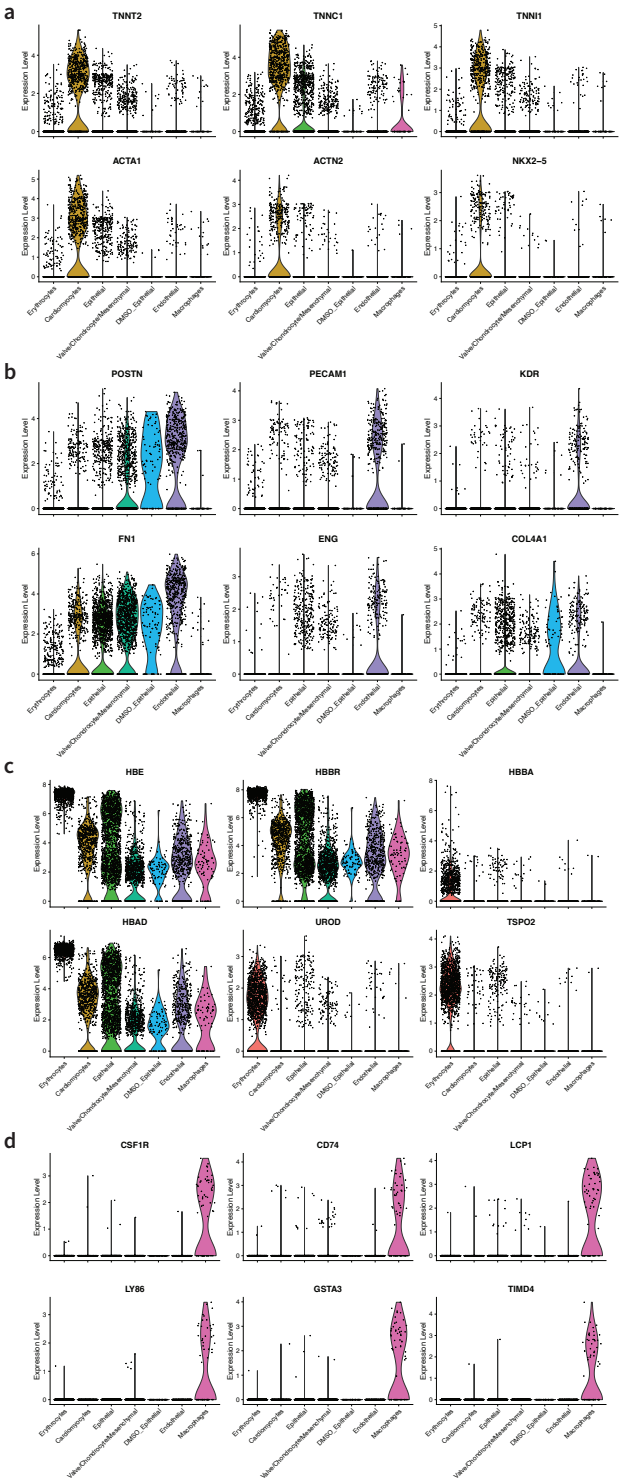


Fig S5.4 Cell type specific expression of genes annotated for **a)** troponin and actinin **b)** basement membrane and angiogenesis **c)** heme biogenesis **d)** macrophage

Chapter 6 Conclusions and Future directions

Lamins are intermediate filament proteins that form a meshwork under the nuclear membrane. In this thesis, we study two major types of Lamins i.e. A-type Lamins and B-type Lamins. Lamin A is coded in the *LMNA* gene present on chromosome 1. Lamin B1 is coded by *LMNB1* gene in chromosome 5 and Lamin B2 is coded in by *LMNB2* in chromosome 19. The two lamins have a largely similar structure except that B-type lamins are farnesylated to the nuclear membrane. There are however reported differences in the phenotype dictated by different lamins, for example, diseases like dilated cardiomyopathy, progeria etc are caused due to Lamin-A mutations but Lamin B1 mutation associated diseases are not widely observed in the human population. Moreover, in contrast to A-type lamins, silencing of B-type lamins leads to apoptotic phenotype in cell culture indicating they are essential after birth (Harborth *et al.*, 2001b). The idea of Lamin B1 being essential is reinforced by studies showing Lamin B1 null mice die soon after birth (Vergnes *et al.*, 2004). Lamin B1 downregulation is also widely accepted as a marker for senescent cells (cells which have exited cell cycle) (Freund, R. M. Laberge, *et al.*, 2012). Lamin A scales with collagen and increasing stiffness of tissue, in contrast to Lamin B1. In developing embryos, Lamin B1 is expressed from an early stage, in contrast Lamin A is almost undetectable and increases as the embryo develops (Cho, Vashisth, *et al.*, 2019). This indicates differences in the roles of the two lamins where Lamin B1 is essential for proliferation, but Lamin A is mechanosensitive. This thesis studies the divergent roles of these two lamins in proliferating tissues of cancer and developing embryos.

6.1 Mechanosensitive Lamin-A protects against DNA damage, telomere attrition and cell cycle arrest

Lamin A is regulated in an acto-myosin stress dependent fashion through phosphorylation during interphase. This perhaps rationalizes many Lamin A mutation diseases being associated with cardiac/skeletal tissues which are stiff and have high acto-myosin stress like dilated cardiomyopathy, muscular dystrophy, progeria. The mechanism through which Lamin A mutation might cause progeria (accelerated aging) is unclear. Interestingly, mice deficient in Ku80 also show

progeroid symptoms and Ku80 plays an important role in telomere maintenance. Progeroid Lamin A is not mechanosensitive and shows decreased interphase phosphorylation (Cho *et al.*, 2018). Chapter 2 shows Ku80 preferentially binds to non-phosphorylated form of Lamin A. Further studies might indicate the role of progeroid Lamin A in sequestration of Ku80 which leads to telomere specific DNA damage.

Chapter 2 shows mis regulation of Lamin A increases probability of nuclear membrane rupture in embryonic heart tissue as well as cancer cells in culture. This results in knockdown of Lamin-A mimicking the phenotype of knockdown of DNA repair factors, leading to accumulation of DNA damage and p53 signaling dependent cell cycle arrest. This also results in DNA damage specific to telomeres. Telomere dysfunction has been associated with cardiomyocyte cell cycle arrest (Aix *et al.*, 2016) and heart failure (M *et al.*, 2017). Perhaps one of the factors instigating the sharp decline in proliferation observed in post-natal hearts is p-53 dependent cell cycle arrest due to accumulated DNA damage in embryonic hearts.

Chapter 5 studies methods to link phenotype to spontaneously arising CNV in lung adenocarcinoma epithelial cells. Interestingly, copy number variation has also been reported in 'normal' cells like liver, brain, donor derived fibroblasts, neurons differentiated from human induced pluripotent stem cells (hiPSC) and post-mortem brains (McConnell *et al.*, 2013; Knouse *et al.*, 2014). There have also been reports of early stage embryonic cells exhibiting aneuploidy which decrease at later stages of embryonic development (MC *et al.*, 2000; C *et al.*, 2007). Heart is the first functional organ to develop in a vertebrate. Embryonic heart nuclei experience acute acto-myosin stress. Chapter 2 shows mis-regulation of mechanosensitive LMNA in embryonic heart cells leads to accumulation of DNA damage and they undergo cell cycle arrest. Future work studying scDNA seq in embryonic hearts with mis regulated LMNA or externally induced DNA damage can depict CNV. InferCNV and UMAP pipeline developed in Chapter 5 can be used to study cell type specific CNV and cell cycle arrest phenotype. DMSO Control hearts in chapter 2 also show basal level of mis localization of DNA repair factors, DNA damage and p21 expression. Perhaps, cells that show cell cycle arrest are also the ones that have CNV. Moreover, DNA damage due to

replication errors (replication stress) play a major role in tumorigenesis (Tomasetti and Vogelstein, 2015; C, L and B, 2017). Chapter 3 shows DNA repair factors are upregulated in proliferating primary tumors compared to adjacent uninvolved. Permanent exit from cell cycle in post-natal hearts may be a mechanism to prevent tumorigenesis given that heart tumors are extremely rare. Hence, even though lack of regenerative capacity in hearts underlies the pathophysiology of myocardial infarction, perhaps it has evolved to prevent tumorigenesis in this tissue that experiences high mechanical stress. This might be something to keep in mind when developing regenerative therapies for the heart.

6.2 Lamin B1 is a marker for proliferation

Activated oncogene Ras induced senescence is achieved by autophagy mediated degradation of Lamin B1(Dou *et al.*, 2015). Perhaps this indicates that Lamin B1 plays an important role in onset of cancer, and autophagy mediated degradation of Lamin B1 has evolved as a mechanism to prevent tumorigenesis through activation of oncogene Ras. Moreover, Lamin B1 is essential for embryonic development(Harborth *et al.*, 2001a), is constitutively expressed early in development (Sangkyun Cho, Vashisth, *et al.*, 2019) and closely associates with chromatin in late mitosis (Gerace and Blobel, 1980). Chapter 3 shows that LMNB1 is upregulated in 15 cancer and predicts significantly poor survival in 8 cancers. Lamin B1 differs from Lamin A besides structural differences in terms of farnesylation, chapter 3 shows that Lamin B1 is regulated by oncogene FOXM1 and acts as a functional marker for cancer. Blood circulating LMNB1 is used as a prognostic marker in Liver Cancer (Sadek Rezk *et al.*, 2015; Abdelghany *et al.*, 2018), given that it also predicts poor survival in 7 other cancer, more expansive studies in other cancers can be conducted. Blood circulating markers are important for early detection and early detection can make a big difference in survival of patient. *LMNB1* scales with mitosis associated genes like *KIF18B, TOP2A, KIF20A, CENPF* in >17 cancers. Moreover, the other farnesylated lamin- Lamin B2 has been shown to play a vital role in karyokinesis (Han *et al.*, 2020). Future work would perhaps study the role of Lamin B1 in mitosis and karyokinesis which involves kinesins, like that of Lamin B2.

6.3 'Universal' scaling laws and dimensional reduction techniques help visualize meta-patterns

Each cancer can have a very specific pathophysiology, indeed it is hard to find a 'cure-all' panacea for cancer, given each type of cancer is a very different disease altogether. As personalized medicine emerges and more and more patients get their genome sequenced, we find that within a cancer type also patients can have a multitude of different malignancies which evolve with time. However, scaling laws in biology like Klieber's power law (KLEIBER, 1947) (Thommen et al., 2019) metabolic rate ~ (body mass)$^{0.75}$ hold true for seemingly disparate entities ranging from the mitochondria to the elephant. Universal scaling laws can also enable prediction of mechanism, for example Hemoglobin binding of oxygen fits a characteristic power law of ~3 at half-saturation (the "Hill coefficient"), prompting Pauling to propose a now well-known molecular mechanism (Pauling, 1935). In chapter 3 we find similar 'universal' scaling laws hold for gene expression in cancer and can help discover novel 'pan-cancer' transcriptional regulators, for example FOXM1 regulates transcription of *LMNB1* (**Fig 3.4**). Specific values of power law exponents are informative. For example, high levels of the *ACTA2* contractile gene set is pro-survival in liver cancer based on (Tumor/Adjacent) data and also for the larger patient sets of tumor-only data except for the one contractile gene with the lowest power: *TAGLN* shows no significance in survival in the tumor-only data (**Fig.S3.7d**). All of the contractile genes scale (strongly) with *ACTA2* exponents >1. In comparison, the *COL1A1* synthetic gene set is pro-survival only in the (Tumor/Adjacent) data and is not significant in nearly all of the larger patient sets of tumor-only data (**Fig.3.7e,f,S3.7d**); and unlike the *ACTA2* exponents, all(except *POSTN*) of the synthetic genes scale (weakly) with *COL1A1* exponents <1 (**Fig.3.6d**), which means *COL1A1* scales super-stoichiometrically with respect to the other genes. A reasonable analogy is that collagen-1 forms the woody bulk of a tree that the other factors assemble onto like bark or leaves on the tree. Regardless of the utility of this analogy, the contractile gene set is overall more sensitive to any changes in *ACTA2.*

In recent experiments(Dingal and Discher, 2015), ACTA2 protein increases dramatically with a strong, switch-like cooperativity exponent of ~3 as a mechano-repressive transcription factor exits the nucleus in mesenchymal cells that spread on stiff collagenous substrates but not on soft

substrates. Importantly, *COL1A1* shows no change when we overexpressed the same transcription factor in cells on rigid plastic, which drives its nuclear accumulation and represses *ACTA2* as expected. Thus specific power law exponents help to indicate strong (*ACTA2*) or weak (*COL1A1*) regulation, extending it seems to data for patient survival.

Machine learning methods for dimensional reduction are important to process big data, especially if the data is sparse. Chapter 5 elaborates methods that can be used to link CNV to phenotype and identify constituent cell type in tissues.

Applications of Scaling-informed Machine Learning (SIML)

In chapter 3 we discover that 'universal' scaling laws and proliferation markers can inform machine learning dimensional reduction methods to form clusters predictive of patient survival.

Scaling laws can also identify genes sets that are useful for cell-type identification in scRNA sequencing, for example, identifying genes that differentiate 'synthetic' CAFs from 'contractile' CAFs.

This thesis combines wet lab experiments and machine learning computational methods to study the role of Lamins in proliferating cells.

BIBLIOGRAPHY

A, S. (2012) 'Telomeres at a glance', *Journal of cell science*. J Cell Sci, 125(Pt 18), pp. 4173–4178. doi: 10.1242/JCS.106831.

A, S. and T, de L. (2012) 'Removal of shelterin reveals the telomere end-protection problem', *Science (New York, N.Y.)*. Science, 336(6081), pp. 593–597. doi: 10.1126/SCIENCE.1218498.

Abdelghany, A. M. *et al.* (2018) 'Using Lamin B1 mRNA for the early diagnosis of hepatocellular carcinoma: a cross-sectional diagnostic accuracy study', *F1000Research*, 7, p. 1339. doi: 10.12688/f1000research.14795.1.

Aiello, K. A. and Alter, O. (2016) 'Platform-Independent Genome-Wide Pattern of DNA Copy-Number Alterations Predicting Astrocytoma Survival and Response to Treatment Revealed by the GSVD Formulated as a Comparative Spectral Decomposition', *PLOS ONE*. Edited by S. D. Peddada, 11(10), p. e0164546. doi: 10.1371/journal.pone.0164546.

Aix, E. *et al.* (2016) 'Postnatal telomere dysfunction induces cardiomyocyte cell-cycle arrest through p21 activation', *The Journal of Cell Biology*. The Rockefeller University Press, 213(5), p. 571. doi: 10.1083/JCB.201510091.

Arts, G. J. *et al.* (1998) 'The role of exportin-t in selective nuclear export of mature tRNAs', *EMBO Journal*. European Molecular Biology Organization, 17(24), pp. 7430–7441. doi: 10.1093/emboj/17.24.7430.

Avery, D. *et al.* (2018) 'Extracellular matrix directs phenotypic heterogeneity of activated fibroblasts', *Matrix Biology*. Elsevier B.V., 67, pp. 90–106. doi: 10.1016/j.matbio.2017.12.003.

B, van S., A, S. and T, de L. (1998) 'TRF2 protects human telomeres from end-to-end fusions', *Cell*. Cell, 92(3), pp. 401–413. doi: 10.1016/S0092-8674(00)80932-0.

Bhat, U. G., Halasi, M. and Gartel, A. L. (2009) 'Thiazole antibiotics target FoxM1 and induce apoptosis in human cancer cells', *PLoS ONE*. PLoS One, 4(5). doi: 10.1371/journal.pone.0005592.

Boulton, S. J. and Jackson, S. P. (1998) 'Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing', *EMBO Journal*, 17(6), pp. 1819–1828. doi: 10.1093/emboj/17.6.1819.

Bradley, M. W. *et al.* (2019) 'GSVD- and tensor GSVD-uncovered patterns of DNA copy-number alterations predict adenocarcinomas survival in general and in response to platinum', *APL Bioengineering*. AIP Publishing, 3(3), p. 036104. doi: 10.1063/1.5099268.

Butler, A. *et al.* (2018) 'Integrating single-cell transcriptomic data across different conditions, technologies, and species', *Nature Biotechnology*. Nature Publishing Group, 36(5), pp. 411–420. doi: 10.1038/nbt.4096.

Buxboim, A. *et al.* (2014) 'Matrix elasticity regulates lamin-A,C phosphorylation and turnover with feedback to actomyosin.', *Current biology : CB*. Elsevier, 24(16), pp. 1909–17. doi: 10.1016/j.cub.2014.07.001.

C, R. *et al.* (2007) 'Impact of chromosomal abnormalities on preimplantation embryo development', *Prenatal diagnosis*. Prenat Diagn, 27(8), pp. 748–756. doi: 10.1002/PD.1773.

C, T., L, L. and B, V. (2017) 'Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention', *Science (New York, N.Y.)*. Science, 355(6331), pp. 1330–1334. doi: 10.1126/SCIENCE.AAF9011.

Capo-chichi, C. D. *et al.* (2011) 'Loss of A-type lamin expression compromises nuclear envelope integrity in breast cancer.', *Chinese journal of cancer*, 30(6), pp. 415–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21627864 (Accessed: 18 February 2019).

Chang, A. C. Y. *et al.* (2018) 'Telomere shortening is a hallmark of genetic cardiomyopathies', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(37), pp. 9276–9281. doi: 10.1073/pnas.1714538115.

Chen, X. *et al.* (2013) 'The Forkhead Transcription Factor FOXM1 Controls Cell Cycle-Dependent Gene Expression through an Atypical Chromatin Binding Mechanism', *Molecular and Cellular Biology*. American Society for Microbiology, 33(2), pp. 227–236. doi: 10.1128/mcb.00881-12.

Cho, S. *et al.* (2018) 'Progerin phosphorylation in interphase is lower and less mechanosensitive than lamin-A,C in iPS-derived mesenchymal stem cells', *Nucleus*, 9. doi: 10.1080/19491034.2018.1460185.

Cho, Sangkyun, Vashisth, M., Abbas, A., Majkut, S., Vogel, K., Xia, Y., Ivanovska, Irena L, *et al.* (2019) 'Mechanosensing by the Lamina Protects against Nuclear Rupture, DNA Damage, and Cell-Cycle Arrest.', *Developmental cell*. Elsevier, 49(6), pp. 920-935.e5. doi: 10.1016/j.devcel.2019.04.020.

Cho, S. *et al.* (2019) 'Mechanosensing by the Lamina Protects against Nuclear Rupture, DNA Damage, and Cell-Cycle Arrest', *Developmental Cell*, 49(6). doi: 10.1016/j.devcel.2019.04.020.

Cho, Sangkyun, Vashisth, M., Abbas, A., Majkut, S., Vogel, K., Xia, Y., Ivanovska, Irena L., *et al.* (2019) 'Mechanosensing by the Lamina Protects against Nuclear Rupture, DNA Damage, and Cell-Cycle Arrest', *Developmental Cell*. Cell Press, 49(6), pp. 920-935.e5. doi: 10.1016/j.devcel.2019.04.020.

Cho, S., Irianto, J. and Discher, D. E. (2017) 'Mechanosensing by the nucleus: From pathways to scaling relationships.', *The Journal of cell biology*. Rockefeller University Press, 216(2), pp. 305–315. doi: 10.1083/jcb.201610042.

Clark, R. A. *et al.* (1982) 'Fibronectin and fibrin provide a provisional matrix for epidermal cell migration during wound reepithelialization.', *The Journal of investigative dermatology*, 79(5), pp. 264–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6752288 (Accessed: 10 June 2019).

Dahl, K. N. *et al.* (2005) 'Power-law rheology of isolated nuclei with deformation mapping of nuclear substructures', *Biophysical Journal*. Elsevier, 89(4), pp. 2855–2864. doi: 10.1529/biophysj.105.062554.

Dasika, G. K. *et al.* (1999) 'DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis', *Oncogene*, 18(55), pp. 7883–7899. doi: 10.1038/sj.onc.1203283.

Davis, C. A. *et al.* (2018) 'The Encyclopedia of DNA elements (ENCODE): Data portal update', *Nucleic Acids Research*. Oxford University Press, 46(D1), pp. D794–D801. doi: 10.1093/nar/gkx1081.

Dechat, T. *et al.* (2010) 'Nuclear Lamins (Review)', *Cold Spring Harbor perspectives in biology*, 2(11), pp. 1–23. doi: 10.1101/cshperspect.a000547.

Denais, C. M. *et al.* (2016) 'Nuclear envelope rupture and repair during cancer cell migration', *Science*, 352(6283), pp. 353–358.

Dingal, P. C. D. P. *et al.* (2015) 'Fractal heterogeneity in minimal matrix models of scars modulates stiff-niche stem-cell responses via nuclear exit of a mechanorepressor', *Nature Materials*. Nature Publishing Group, 14(9), pp. 951–960. doi: 10.1038/nmat4350.

Dingal, P. C. D. P. and Discher, D. E. (2015) 'Systems mechanobiology: Tension-inhibited protein turnover is sufficient to physically control gene circuits', *Biophysical Journal*. Biophysical Society, 107(11), pp. 2734–2743. doi: 10.1016/j.bpj.2014.10.042.

Dittmer, T. A. and Misteli, T. (2011) 'The lamin protein family.', *Genome biology*. BioMed Central, 12(5), p. 222. doi: 10.1186/gb-2011-12-5-222.

Dou, Z. *et al.* (2015) 'Autophagy mediates degradation of nuclear lamina', *Nature*. Nature Publishing Group, 527(7576), pp. 105–109. doi: 10.1038/nature15548.

Dupont, S. *et al.* (2011) 'Role of YAP/TAZ in mechanotransduction', *Nature*. Nature Publishing Group, 474(7350), pp. 179–183. doi: 10.1038/nature10137.

Eggan, K. *et al.* (2002) 'Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation', *Nature Biotechnology*. Nat Biotechnol, 20(5), pp. 455–459. doi: 10.1038/nbt0502-455.

Engler, A. J. *et al.* (2006) 'Matrix Elasticity Directs Stem Cell Lineage Specification', *Cell*, 126(4), pp. 677–689. doi: 10.1016/j.cell.2006.06.044.

Fink, L. S. *et al.* (2010) 'Ku80 facilitates chromatin binding of the telomere binding protein, TRF2', *Cell Cycle*. Taylor and Francis Inc., 9(18), pp. 3798–3806. doi: 10.4161/cc.9.18.13129.

Freund, A., Laberge, R.-M., *et al.* (2012) 'Lamin B1 loss is a senescence-associated biomarker'. doi: 10.1091/mbc.E11-10-0884.

Freund, A., Laberge, R. M., *et al.* (2012) 'Lamin B1 loss is a senescence-associated biomarker', *Molecular Biology of the Cell*. American Society for Cell Biology, 23(11), pp. 2066–2075. doi: 10.1091/mbc.E11-10-0884.

G, G. *et al.* (2007) 'The lethality of Ku86 (XRCC5) loss-of-function mutations in human cells is independent of p53 (TP53)', *Radiation research*. Radiat Res, 167(1), pp. 66–79. doi: 10.1667/RR0692.1.

Garside, V. C. *et al.* (2015) 'SOX9 modulates the expression of key transcription factors required for heart valve development', *Development (Cambridge)*. Company of Biologists Ltd, 142(24), pp. 4340–4350. doi: 10.1242/dev.125252.

Gelse, K., Pöschl, E. and Aigner, T. (2003) 'Collagens--structure, function, and biosynthesis.', *Advanced drug delivery reviews*, 55(12), pp. 1531–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14623400 (Accessed: 26 February 2019).

Gerace, L. and Blobel, G. (1980) 'The nuclear envelope lamina is reversibly depolymerized during mitosis.', *Cell*. Elsevier, 19(1), pp. 277–87. doi: 10.1016/0092-8674(80)90409-2.

Goldman, A. E. *et al.* (1992) 'Pathway of incorporation of microinjected lamin A into the nuclear envelope', *Journal of Cell Biology*. J Cell Biol, 119(4), pp. 725–735. doi: 10.1083/jcb.119.4.725.

Gormally, M. V. *et al.* (2014) 'Suppression of the FOXM1 transcriptional programme via novel small molecule inhibition', *Nature Communications*. Nature Publishing Group, 5(1), pp. 1–11. doi: 10.1038/ncomms6165.

Grant, G. D. *et al.* (2013) 'Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors', *Molecular Biology of the Cell*. The American Society for Cell Biology , 24(23), pp. 3634–3650. doi: 10.1091/mbc.E13-05-0264.

Gruenbaum, Y. and Aebi, U. (2014) 'Intermediate filaments: a dynamic network that controls cell mechanics.', *F1000prime reports*. Faculty of 1000 Ltd, 6, p. 54. doi: 10.12703/P6-54.

Guilak, F., Tedrow, J. R. and Burgkart, R. (2000) 'Viscoelastic Properties of the Cell Nucleus', *Biochemical and Biophysical Research Communications*. Academic Press, 269(3), pp. 781–786. doi: 10.1006/BBRC.2000.2360.

Guilluy, C. *et al.* (2014) 'Isolated nuclei adapt to force and reveal a mechanotransduction pathway within the nucleus', *Nat Cell Biol*, 16(4), pp. 376–381. doi: 10.1038/ncb2927.

H, W. *et al.* (2005) 'DNA ligase III as a candidate component of backup pathways of nonhomologous end joining', *Cancer research*. Cancer Res, 65(10), pp. 4020–4030. doi: 10.1158/0008-5472.CAN-04-3055.

Han, L. *et al.* (2020) 'Lamin B2 Levels Regulate Polyploidization of Cardiomyocyte Nuclei and Myocardial Regeneration Article Lamin B2 Levels Regulate Polyploidization of Cardiomyocyte Nuclei and Myocardial Regeneration', *Developmental Cell*. Cell Press, pp. 1–18. doi: 10.1016/j.devcel.2020.01.030.

Harada, T. *et al.* (2014a) 'Nuclear lamin stiffness is a barrier to 3D migration, but softness can limit survival', *The Journal of Cell Biology*, 204(5), pp. 669–682. doi: 10.1083/jcb.201308029.

Harada, T. *et al.* (2014b) 'Nuclear lamin stiffness is a barrier to 3D migration, but softness can limit survival', *J Cell Biol*. Rockefeller University Press, 204(5), pp. 669–682. doi: 10.1083/JCB.201308029.

Harborth, J. *et al.* (2001a) 'Identification of essential genes in cultured mammalian cells using small interfering RNAs.', *Journal of cell science*, 114(Pt 24), pp. 4557–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11792820 (Accessed: 8 August 2019).

Harborth, J. *et al.* (2001b) 'Identification of essential genes in cultured mammalian cells using small interfering RNAs.', *Journal of cell science*, 114(Pt 24), pp. 4557–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11792820 (Accessed: 31 July 2019).

Hayflick, L. and Moorhead, P. S. (1961) 'The serial cultivation of human diploid cell strains', *Experimental Cell Research*. Exp Cell Res, 25(3), pp. 585–621. doi: 10.1016/0014-4827(61)90192-6.

Hennekes, H. and Nigg, E. A. (1994) 'The role of isoprenylation in membrane attachment of nuclear lamins', *Journal of Cell Science*, 107, pp. 1019–1029.

Hernandez, L. *et al.* (2010) 'Functional Coupling between the Extracellular Matrix and Nuclear Lamina by Wnt Signaling in Progeria', *Developmental Cell*, 19(3), pp. 413–425. doi: 10.1016/j.devcel.2010.08.013.

Higo, T. *et al.* (2017) 'DNA single-strand break-induced DNA damage response causes heart failure', *Nature Communications*. Nature Publishing Group, 8. doi: 10.1038/ncomms15104.

Ho, C. Y. *et al.* (2013) 'Lamin A/C and emerin regulate MKL1–SRF activity by modulating actin dynamics', *Nature*. Nature Publishing Group, 497(7450), pp. 507–511. doi: 10.1038/nature12105.

Hofheinz, R.-D. *et al.* (2003) 'Stromal Antigen Targeting by a Humanised Monoclonal Antibody: An Early Phase II Trial of Sibrotuzumab in Patients with Metastatic Colorectal Cancer', *Oncology Research and Treatment*. Karger Publishers, 26(1), pp. 44–48. doi: 10.1159/000069863.

Hsu, H.-L. *et al.* (2000) 'Ku acts in a unique way at the mammalian telomere to prevent end joining', *Genes & Development*. Cold Spring Harbor Laboratory Press, 14(22), pp. 2807–2812. doi: 10.1101/GAD.844000.

Huang, X., Halicka, H. D. and Darzynkiewicz, Z. (2004) 'Detection of histone H2AX phosphorylation on Ser-139 as an indicator of DNA damage (DNA double-strand breaks).', *Current protocols in cytometry / editorial board, J. Paul Robinson, managing editor ... [et al.]*. Curr Protoc Cytom, Chapter 7. doi: 10.1002/0471142956.cy0727s30.

I, J., P, M. and MA, B. (2004) 'Role of human Ku86 in telomere length maintenance and telomere capping', *Cancer research*. Cancer Res, 64(20), pp. 7271–7278. doi: 10.1158/0008-5472.CAN-04-1381.

Irianto, J., Xia, Y., *et al.* (2016) 'As a nucleus enters a small pore, chromatin stretches and maintains integrity, even with DNA breaks', *Biophysical Journal*. Biophysical Society, 112(3), pp. 446–449. doi: 10.1016/j.bpj.2016.09.047.

Irianto, J., Pfeifer, Charlotte R, *et al.* (2016) 'Nuclear constriction segregates mobile nuclear proteins away from chromatin.', *Molecular biology of the cell*, 27(25), pp. 4011–4020. doi: 10.1091/mbc.E16-06-0428.

Irianto, J., Pfeifer, Charlotte R., *et al.* (2016) 'Nuclear Lamins in Cancer', *Cellular and Molecular Bioengineering*. Springer New York LLC, 9(2), pp. 258–267. doi: 10.1007/s12195-016-0437-8.

Irianto, J. *et al.* (2017) 'DNA damage follows repair factor depletion and portends genome

variation in cancer cells after pore migration', *Current Biology*. Elsevier Ltd., 27(2), pp. 210–223. doi: 10.1016/j.cub.2016.11.049.

Jafarpour, F. *et al.* (2018) 'Bridging the Timescales of Single-Cell and Population Dynamics', *Physical Review X*. American Physical Society, 8(2), p. 021007. doi: 10.1103/PhysRevX.8.021007.

K, M. *et al.* (2004) 'Regulation of telomere length and suppression of genomic instability in human somatic cells by Ku86', *Molecular and cellular biology*. Mol Cell Biol, 24(11), pp. 5050–5059. doi: 10.1128/MCB.24.11.5050-5059.2004.

Kahlem, P. *et al.* (2004) 'Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of Down syndrome', *Genome Research*. Genome Res, 14(7), pp. 1258–1267. doi: 10.1101/gr.1951304.

Kanellos, G. *et al.* (2015) 'ADF and Cofilin1 Control Actin Stress Fibers, Nuclear Integrity, and Cell Survival', *Cell Reports*. Elsevier B.V., 13(9), pp. 1949–1964. doi: 10.1016/j.celrep.2015.10.056.

Kaspi, E. *et al.* (2017) 'Low lamin A expression in lung adenocarcinoma cells from pleural effusions is a pejorative factor associated with high number of metastatic sites and poor Performance status', *PLOS ONE*. Edited by O. Ogunwobi, 12(8), p. e0183136. doi: 10.1371/journal.pone.0183136.

Kilvaer, T. K. *et al.* (2015) 'Cancer associated fibroblasts in stage I-IIIA NSCLC: Prognostic impact and their correlations with tumor molecular markers', *PLoS ONE*. Public Library of Science, 10(8), p. e0134965. doi: 10.1371/journal.pone.0134965.

KLEIBER, M. (1947) 'Body size and metabolic rate', *Physiological reviews*, 27(4), pp. 511–541. doi: 10.1152/physrev.1947.27.4.511.

Knouse, K. A. *et al.* (2014) 'Single cell sequencing reveals low levels of aneuploidy across mammalian tissues', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(37), pp. 13409–13414. doi: 10.1073/pnas.1415287111.

Knudson, A. G. (1971) 'Mutation and cancer: statistical study of retinoblastoma.', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 68(4), pp. 820–823. doi: 10.1073/pnas.68.4.820.

Kochanek, K. D., Xu, J. and Arias, E. (2019) *Key findings Data from the National Vital Statistics System How long can we expect to live?* Available at: https://www.cdc.gov/nchs/products/index.htm. (Accessed: 1 July 2021).

Koike, M., Shiomi, T. and Koike, A. (2001) 'Dimerization and Nuclear Localization of Ku Proteins', *Journal of Biological Chemistry*. J Biol Chem, 276(14), pp. 11167–11173. doi: 10.1074/jbc.M010902200.

Kurosaka, S. and Kashina, A. (2008) 'Cell biology of embryonic migration', *Birth Defects Research Part C: Embryo Today: Reviews*, 84(2), pp. 102–122. doi: 10.1002/bdrc.20125.

de la Rosa, J. *et al.* (2013) 'Prelamin A causes progeria through cell-extrinsic mechanisms and prevents cancer invasion.', *Nature communications*. Europe PMC Funders, 4, p. 2268. doi: 10.1038/ncomms3268.

Lamalice, L., Le Boeuf, F. and Huot, J. (2007) 'Endothelial Cell Migration During Angiogenesis', *Circulation Research*, 100(6), pp. 782–794. doi: 10.1161/01.RES.0000259593.07661.1e.

Laoukili, J. *et al.* (2005) 'FoxM1 is required for execution of the mitotic programme and chromosome stability', *Nature Cell Biology*. Nature Publishing Group, 7(2), pp. 126–136. doi: 10.1038/ncb1217.

Laoukili, J. *et al.* (2008) 'FoxM1 is degraded at mitotic exit in a Cdh1-dependent manner', *Cell Cycle*. Taylor and Francis Inc., 7(17), pp. 2720–2726. doi: 10.4161/cc.7.17.6580.

Levental, K. R. *et al.* (2009) 'Matrix crosslinking forces tumor progression by enhancing integrin signaling.', *Cell*. NIH Public Access, 139(5), pp. 891–906. doi: 10.1016/j.cell.2009.10.027.

Li, F. *et al.* (1996) 'Rapid transition of cardiac myocytes from hyperplasia to hypertrophy during postnatal development', *Journal of Molecular and Cellular Cardiology*. Academic Press, 28(8), pp. 1737–1746. doi: 10.1006/jmcc.1996.0163.

Lin, R. *et al.* (2006) 'Oxidative DNA damage and DNA mismatch repair pathway play an important role in failing human myocardium', *Journal of Cardiothoracic-Renal Research*, 1(1), pp. 41–49. doi: 10.1016/j.jccr.2005.11.004.

Liotta, L. A., Steeg, P. S. and Stetler-Stevenson, W. G. (1991) 'Cancer metastasis and angiogenesis: An imbalance of positive and negative regulation', *Cell*, 64(2), pp. 327–336. doi: 10.1016/0092-8674(91)90642-C.

M, A. *et al.* (2006) 'Involvement of polynucleotide kinase in a poly(ADP-ribose) polymerase-1-dependent DNA double-strand breaks rejoining pathway', *Journal of molecular biology*. J Mol Biol, 356(2), pp. 257–265. doi: 10.1016/J.JMB.2005.11.028.

M, M. *et al.* (2021) 'Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis', *Nature communications*. Nat Commun, 12(1). doi: 10.1038/S41467-021-21892-Z.

M, S.-S. *et al.* (2017) 'Cardiomyocyte-Specific Telomere Shortening is a Distinct Signature of Heart Failure in Humans', *Journal of the American Heart Association*. J Am Heart Assoc, 6(9). doi: 10.1161/JAHA.116.005086.

Ma, L. *et al.* (2019) 'Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer', *Cancer Cell*. Cell Press, 36(4), pp. 418-430.e6. doi: 10.1016/j.ccell.2019.08.007.

Maciejowski, J. *et al.* (2015) 'Chromothripsis and Kataegis Induced by Telomere Crisis', *Cell*. Cell Press, 163(7), pp. 1641–1654. doi: 10.1016/j.cell.2015.11.054.

Major, M. L., Lepe, R. and Costa, R. H. (2004) 'Forkhead Box M1B Transcriptional Activity Requires Binding of Cdk-Cyclin Complexes for Phosphorylation-Dependent Recruitment of p300/CBP Coactivators', *Molecular and Cellular Biology*. American Society for Microbiology, 24(7), pp. 2649–2661. doi: 10.1128/mcb.24.7.2649-2661.2004.

Marti, T. M. *et al.* (2006) 'H2AX phosphorylation within the G1 phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 103(26), pp. 9891–9896. doi: 10.1073/PNAS.0603779103.

MC, M. *et al.* (2000) 'Chromosome mosaicism in day 3 aneuploid embryos that develop to morphologically normal blastocysts in vitro', *Human reproduction (Oxford, England)*. Hum Reprod, 15(8), pp. 1781–1786. doi: 10.1093/HUMREP/15.8.1781.

McConnell, M. J. *et al.* (2013) 'Mosaic copy number variation in human neurons', *Science*. American Association for the Advancement of Science, 342(6158), pp. 632–637. doi: 10.1126/science.1243472.

McInnes, L., Healy, J. and Melville, J. (2018) 'UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction'. Available at: http://arxiv.org/abs/1802.03426 (Accessed: 4 June 2021).

Melnikova, L., Biessmann, H. and Georgiev, P. (2005) 'The Ku protein complex is involved in length regulation of Drosophila telomeres', *Genetics*, 170(1), pp. 221–235. doi: 10.1534/genetics.104.034538.

Mendez, M. G., Kojima, S.-I. and Goldman, R. D. (2010) 'Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition', *The FASEB Journal*. The Federation of American Societies for Experimental Biology, 24(6), p. 1838. doi: 10.1096/FJ.09-151639.

Meng, Z. *et al.* (2018) 'RAP2 mediates mechanoresponses of the Hippo pathway', *Nature*. Nature Publishing Group, 560(7720), pp. 655–660. doi: 10.1038/s41586-018-0444-0.

Mermel, C. H. *et al.* (2011) 'GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers', *Genome Biology*. BioMed Central, 12(4), p. R41. doi: 10.1186/gb-2011-12-4-r41.

Miralles, F. *et al.* (2003) 'Actin Dynamics Control SRF Activity by Regulation of Its Coactivator MAL', *Cell*. Cell Press, 113(3), pp. 329–342. doi: 10.1016/S0092-8674(03)00278-2.

Moir, R. D., Spann, T. P., *et al.* (2000) *Disruption of Nuclear Lamin Organization Blocks the Elongation Phase of DNA Replication*, *The Journal of Cell Biology*. Available at: http://www.jcb.org (Accessed: 21 February 2019).

Moir, R. D., Yoon, M., *et al.* (2000) *Nuclear Lamins A and B1: Different Pathways of Assembly during Nuclear Envelope Formation in Living Cells*, *The Journal of Cell Biology*. Available at: http://www.jcb.org/cgi/content/full/151/6/1155 (Accessed: 21 February 2019).

Moir, R. D., Montag-Lowy, M. and Goldman, R. D. (1994) 'Dynamic properties of nuclear lamins: lamin B is associated with sites of DNA replication.', *The Journal of cell biology*. Rockefeller University Press, 125(6), pp. 1201–12. doi: 10.1083/JCB.125.6.1201.

Mondal, N. K. *et al.* (2013) 'Oxidative stress, DNA damage and repair in heart failure patients after implantation of continuous flow left ventricular assist devices', *International Journal of Medical Sciences*, 10(7), pp. 883–893. doi: 10.7150/ijms.6219.

Nielsen, C. F. *et al.* (2020) 'Topoisomerase IIa is essential for maintenance of mitotic chromosome structure', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 117(22), pp. 12131–12142. doi: 10.1073/pnas.2001760117.

Oh, H. *et al.* (2003) 'Telomere attrition and Chk2 activation in human heart failure', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 100(9), pp. 5378–5383. doi: 10.1073/pnas.0836098100.

Öhlund, D. *et al.* (2017) 'Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer', *The Journal of experimental medicine*. The Rockefeller University Press, 214(3), pp. 579–596. doi: 10.1084/jem.20162024.

Özdemir, B. C. *et al.* (2014) 'Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival', *Cancer Cell*, 25(6), pp. 719–734. doi: 10.1016/j.ccr.2014.04.005.

PA, M.-G. *et al.* (2015) 'Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination', *Nature*. Nature, 518(7538), pp. 254–257. doi: 10.1038/NATURE14157.

Pajerowski, J. D. *et al.* (2007) 'Physical plasticity of the nucleus in stem cell differentiation', *PNAS*, 104, pp. 15619–15624.

Palozola, K. C. *et al.* (2017) 'Mitotic transcription and waves of gene reactivation during mitotic exit', *Science*. American Association for the Advancement of Science, 358(6359), pp. 119–122. doi: 10.1126/science.aal4671.

Park, H. J. *et al.* (2008) 'Anaphase-Promoting Complex/Cyclosome-Cdh1-Mediated Proteolysis of the Forkhead Box M1 Transcription Factor Is Critical for Regulated Entry into S Phase', *Molecular and Cellular Biology*. American Society for Microbiology, 28(17), pp. 5162–5171. doi: 10.1128/mcb.00387-08.

Patel, A. P. *et al.* (2014) 'Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma', *Science*. American Association for the Advancement of Science, 344(6190), pp. 1396–1401. doi: 10.1126/science.1254257.

Pauling, L. (1935) 'The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 21(4), p. 186. doi: 10.1073/PNAS.21.4.186.

Peter, A. and Stick, R. (2012) 'Evolution of the lamin protein family: What introns can tell', *Nucleus*. Taylor and Francis Inc., pp. 44–59. doi: 10.4161/nucl.18927.

Pfau, S. J. and Amon, A. (2012) 'Chromosomal instability and aneuploidy in cancer: from yeast to man.', *EMBO reports*. European Molecular Biology Organization, 13(6), pp. 515–27. doi: 10.1038/embor.2012.65.

Pfeifer, C. R. *et al.* (2018) 'Constricted migration increases DNA damage and independently represses cell cycle', *Molecular Biology of the Cell*. Edited by V. M. Weaver, 29(16), pp. 1948–1962. doi: 10.1091/mbc.E18-02-0079.

Pinter, M. *et al.* (2016) 'Cancer and liver cirrhosis: implications on prognosis and management.', *ESMO open*. BMJ Publishing Group, 1(2), p. e000042. doi: 10.1136/esmoopen-2016-000042.

Porrello, E. R. *et al.* (2011) 'Transient regenerative potential of the neonatal mouse heart', *Science*. Science, 331(6020), pp. 1078–1080. doi: 10.1126/science.1200708.

Porrello, E. R. *et al.* (2013) 'Regulation of neonatal and adult mammalian heart regeneration by the miR-15 family', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 110(1), pp. 187–192. doi: 10.1073/pnas.1208863110.

POWELL, E. O. (1956) 'Growth Rate and Generation Time of Bacteria, with Special Reference to Continuous Culture', *Journal of General Microbiology*. Microbiology Society, 15(3), pp. 492–511. doi: 10.1099/00221287-15-3-492.

Puente, B. N. *et al.* (2014) 'The oxygen-rich postnatal environment induces cardiomyocyte cell-cycle arrest through DNA damage response', *Cell*. Elsevier B.V., 157(3), pp. 565–579. doi: 10.1016/j.cell.2014.03.032.

Qayyum, A. *et al.* (2019) 'Immunotherapy response evaluation with magnetic resonance elastography (MRE) in advanced HCC', *Journal for ImmunoTherapy of Cancer*, 7(1), p. 329. doi: 10.1186/s40425-019-0766-y.

Qian, H. *et al.* (2018) 'The lncRNA MIR4435-2HG promotes lung cancer progression by activating β-catenin signalling', *Journal of Molecular Medicine 2018 96:8*. Springer, 96(8), pp. 753–764. doi: 10.1007/S00109-018-1654-5.

R, R. *et al.* (2010) 'The function of classical and alternative non-homologous end-joining

pathways in the fusion of dysfunctional telomeres', *The EMBO journal*. EMBO J, 29(15), pp. 2598–2610. doi: 10.1038/EMBOJ.2010.142.

Raab, M. *et al.* (2016) 'ESCRT III repairs nuclear envelope ruptures during cell migration to limit DNA damage and cell death.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 352(6283), pp. 359–62. doi: 10.1126/science.aad7611.

Radhakrishnan, S. K. *et al.* (2006) 'Identification of a chemical inhibitor of the oncogenic transcription factor forkhead box M1', *Cancer Research*. Cancer Res, 66(19), pp. 9731–9735. doi: 10.1158/0008-5472.CAN-06-1576.

Reiling, E. *et al.* (2014) 'The progeroid phenotype of Ku80 deficiency Is dominant over DNA-PK CS deficiency', *PLoS ONE*. Public Library of Science, 9(4). doi: 10.1371/journal.pone.0093568.

Reon, B. J. *et al.* (2018) 'LINC00152 Promotes Invasion through a 3′-Hairpin Structure and Associates with Prognosis in Glioblastoma', *Molecular Cancer Research*. American Association for Cancer Research, 16(10), pp. 1470–1482. doi: 10.1158/1541-7786.MCR-18-0322.

S, P. and SJ, B. (2014) 'Double-strand break repair: 53BP1 comes into focus', *Nature reviews. Molecular cell biology*. Nat Rev Mol Cell Biol, 15(1), pp. 7–18. doi: 10.1038/NRM3719.

Sadek Rezk, N. *et al.* (2015) *Circulating mRNA Lamin B1 for Detection of Early Stages of Hepatocellular Carcinoma*. Available at: http://ejgg.journals.ekb.eg/article_5325_4a919360c03c41a879f41b2acd4766db.pdf (Accessed: 6 February 2019).

Sanders, D. A. *et al.* (2015) 'FOXM1 binds directly to non-consensus sequences in the human genome', *Genome Biology*. BioMed Central Ltd., 16(1), p. 130. doi: 10.1186/s13059-015-0696-z.

Sankaranarayanan, P. *et al.* (2015) 'Tensor GSVD of Patient- and Platform-Matched Tumor and Normal DNA Copy-Number Profiles Uncovers Chromosome Arm-Wide Patterns of Tumor-Exclusive Platform-Consistent Alterations Encoding for Cell Transformation and Predicting Ovarian Cancer Survival', *PLOS ONE*. Edited by J. D. Hoheisel, 10(4), p. e0121396. doi: 10.1371/journal.pone.0121396.

Sawada, Y. *et al.* (2006) 'Force Sensing by Mechanical Extension of the Src Family Kinase Substrate p130Cas', *Cell*. Elsevier B.V., 127(5), pp. 1015–1026. doi: 10.1016/j.cell.2006.09.044.

Shah, P. P. *et al.* (2013) 'Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape.', *Genes & development*. Cold Spring Harbor Laboratory Press, 27(16), pp. 1787–99. doi: 10.1101/gad.223834.113.

Sharifi-Sanjani, M. *et al.* (2017) 'Cardiomyocyte-specific telomere shortening is a distinct signature of heart failure in humans', *Journal of the American Heart Association*. John Wiley and Sons Inc., 6(9). doi: 10.1161/JAHA.116.005086.

Shimi, T. *et al.* (2008) 'The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription.', *Genes & development*. Cold Spring Harbor Laboratory Press, 22(24), pp. 3409–21. doi: 10.1101/gad.1735208.

Shimi, T. *et al.* (2011) 'The role of nuclear lamin B1 in cell proliferation and senescence.', *Genes & development*. Cold Spring Harbor Laboratory Press, 25(24), pp. 2579–93. doi: 10.1101/gad.179515.111.

Simsek, D. *et al.* (2011) 'DNA Ligase III Promotes Alternative Nonhomologous End-Joining during Chromosomal Translocation Formation', *PLOS Genetics*. Public Library of Science, 7(6), p. e1002080. doi: 10.1371/JOURNAL.PGEN.1002080.

Simsek, D. and Jasin, M. (2010) 'Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4/ligase IV during chromosomal translocation formation', *Nature structural & molecular biology*. NIH Public Access, 17(4), p. 410. doi: 10.1038/NSMB.1773.

Singla, S. *et al.* (2020) 'Autophagy-mediated apoptosis eliminates aneuploid cells in a mouse model of chromosome mosaicism', *Nature Communications*. Nature Research, 11(1), pp. 1–15. doi: 10.1038/s41467-020-16796-3.

Sun, S. *et al.* (2010) 'Circulating Lamin B1 (LMNB1) Biomarker Detects Early Stages of Liver Cancer in Patients', *Journal of Proteome Research*, 9(1), pp. 70–78. doi: 10.1021/pr9002118.

Swift, J. *et al.* (2013) 'Nuclear Lamin-A Scales with Tissue Stiffness and Enhances Matrix-Directed Differentiation', *Science*, 341(6149), pp. 1240104–1240104. doi: 10.1126/science.1240104.

Taimen, P. *et al.* (2009) 'A progeria mutation reveals functions for lamin A in nuclear assembly, architecture, and chromosome organization', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 106(49), pp. 20788–20793. doi: 10.1073/PNAS.0911895106.

Takaki, T. *et al.* (2017) 'Actomyosin drives cancer cell nuclear dysmorphia and threatens genome stability', *Nature Communications*. Nature Publishing Group, 8(1), pp. 1–13. doi: 10.1038/ncomms16013.

Takeda, A. A. S. *et al.* (2011) 'Structural basis of importin-α-mediated nuclear transport for Ku70 and Ku80', *Journal of Molecular Biology*. Academic Press, 412(2), pp. 226–234. doi: 10.1016/j.jmb.2011.07.038.

Tchou, J. *et al.* (2013) 'Fibroblast activation protein expression by stromal cells and tumor-associated macrophages in human breast cancer', *Human Pathology*. W.B. Saunders, 44(11), pp. 2549–2557. doi: 10.1016/j.humpath.2013.06.016.

Thommen, A. *et al.* (2019) 'Body size-dependent energy storage causes Kleiber's law scaling of the metabolic rate in planarians', *eLife*. NLM (Medline), 8. doi: 10.7554/eLife.38187.

Tichy, E. D. *et al.* (2017) 'Single Stem Cell Imaging and Analysis Reveals Telomere Length Differences in Diseased Human and Mouse Skeletal Muscles.', *Stem cell reports*. Elsevier, 9(4), pp. 1328–1341. doi: 10.1016/j.stemcr.2017.08.003.

Tirosh, I. *et al.* (2016) 'Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq', *Science*, 352(6282), pp. 189–196. doi: 10.1126/science.aad0501.

Tomasetti, C. and Vogelstein, B. (2015) 'Variation in cancer risk among tissues can be explained by the number of stem cell divisions', *Science (New York, N.Y.)*. NIH Public Access, 347(6217), p. 78. doi: 10.1126/SCIENCE.1260825.

Torres, E. M. *et al.* (2007) 'Effects of aneuploidy on cellular physiology and cell division in haploid yeast', *Science*. Science, 317(5840), pp. 916–924. doi: 10.1126/science.1142210.

Turgay, Y. *et al.* (2017) 'The molecular architecture of lamins in somatic cells', *Nature*, 543(7644), pp. 261–264. doi: 10.1038/nature21382.

Ungricht, R. and Kutay, U. (2017) 'Mechanisms and functions of nuclear envelope remodelling', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18, pp. 229–245. doi: 10.1038/nrm.2016.153.

Vergnes, L. *et al.* (2004) 'Lamin B1 is required for mouse development and nuclear integrity', *Proceedings of the National Academy of Sciences of the United States of America*, 101(28), pp. 10428–10433. doi: 10.1073/pnas.0401424101.

Vogel, H. *et al.* (1999) 'Deletion of Ku86 causes early onset of senescence in mice', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(19), pp. 10770–10775. doi: 10.1073/pnas.96.19.10770.

Wagstaff, K. M. *et al.* (2012) 'Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus', *Biochemical Journal*. Portland Press Ltd, 443(3), pp. 851–856. doi: 10.1042/BJ20120150.

Wang, H. X. *et al.* (2020) 'Thymic Epithelial Cells Contribute to Thymopoiesis and T Cell Development', *Frontiers in Immunology*. Frontiers Media S.A., p. 3099. doi: 10.3389/fimmu.2019.03099.

Wang, Y., Ghosh, G. and Hendrickson, E. A. (2009) 'From the Cover: Ku86 represses lethal telomere deletion events in human somatic cells', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(30), p. 12430. doi: 10.1073/PNAS.0903362106.

Weigel, D. and Jäckle, H. (1990) 'The fork head domain: A novel DNA binding motif of eukaryotic transcription factors?', *Cell*. Elsevier, pp. 455–456. doi: 10.1016/0092-8674(90)90439-L.

Williams, B. R. *et al.* (2008) 'Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells', *Science*. Science, 322(5902), pp. 703–709. doi: 10.1126/science.1160058.

X, L. *et al.* (2018) 'LncRNA LINC00152 promoted glioblastoma progression through targeting the miR-107 expression', *Environmental science and pollution research international*. Environ Sci Pollut Res Int, 25(18), pp. 17674–17681. doi: 10.1007/S11356-018-1784-X.

Xavier, E. C. (2012) 'A note on a Maximum k-Subset Intersection problem', *Information Processing Letters*. Elsevier, 112(12), pp. 471–472. doi: 10.1016/j.ipl.2012.03.007.

Xia, Y., Ivanovska, I. L., *et al.* (2018) 'Nuclear rupture at sites of high curvature compromises retention of DNA repair factors', *The Journal of Cell Biology*, 217(11), pp. 3796–3808. doi: 10.1083/jcb.201711161.

Xia, Y., Pfeifer, C. R., *et al.* (2018) 'Rescue of DNA damage in cells after constricted migration reveals bimodal mechano-regulation of cell cycle', *bioRxiv*. Cold Spring Harbor Laboratory, p. 508200. doi: 10.1101/508200.

Xia, Y. *et al.* (2019) 'Rescue of DNA damage after constricted migration reveals a mechano-regulated threshold for cell cycle', *Journal of Cell Biology*. Rockefeller University Press, 218(8), pp. 2542–2563. doi: 10.1083/JCB.201811100.

Yang, Y., Leone, L. M. and Kaufman, L. J. (2009) 'Elastic Moduli of Collagen Gels Can Be Predicted from Two-Dimensional Confocal Microscopy', *Biophysical Journal*, 97(7), pp. 2051–2060. doi: 10.1016/j.bpj.2009.07.035.

Yue, S. (2017) 'Age-associated loss of lamin-B1 in thymic epithelial cells disrupts the thymic microenvironment for thymopoiesis and leads to thymic involution', *The Journal of Immunology*, 198(1 Supplement).

Zhang, M. J., Ntranos, V. and Tse, D. (2020) 'Determining sequencing depth in a single-cell RNA-seq experiment', *Nature Communications*. Nature Research, 11(1), pp. 1–11. doi: 10.1038/s41467-020-14482-y.